

À ceux que j'aime et ceux qui m'aiment

Remerciements

Nous tenons tout d'abord à remercier **ALLAH** le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

La réalisation de ce travail d'initiation à la recherche est une expérience passionnante, qui n'aurait pas été possible sans le soutien de personnes à qui je souhaite exprimer ici mes sincères remerciements.

Une pensée émue va à notre cher enseignant, feu Monsieur **Bénamar Chouaf**, cet éminent Professeur qui nous a fait profiter de son encadrement et son riche savoir. Je lui serai à jamais reconnaissante de m'avoir donné l'opportunité de faire partie de l'école doctoral qu'il a ouvert en 2009.

Nous tenons à remercier très chaleureusement Monsieur **Attouch Mohammed Kadi** qui nous a permis de bénéficier de son encadrement. Les conseils qu'il nous a prodigués, la patience, la confiance qu'il nous a témoignés ont été déterminants dans la réalisation de notre travail de recherche, ainsi que pour sa grande disponibilité, sa rigueur scientifique, son enthousiasme et ses précieux conseils qui m'ont permis de travailler dans les meilleures conditions.

Mes sincères remerciements et ma gratitude vont aussi à Monsieur **TEBBOUNE Fethallah Ouahabi** pour avoir accepté de juger ce travail et d'en présider le jury de soutenance. Que vous soyez assuré de mon entière reconnaissance.

Je remercie Monsieur le Professeur **LAKSACI Ali** d'avoir accepté de rapporter cette thèse. Votre aide et vos remarques pertinentes m'ont beaucoup aidé à améliorer la qualité de ce travail. Soyez assurés de mon estime et de ma profonde gratitude.

Merci également à Monsieur **CHIKER EL MEZOUAR Zouaoui** (Université Béchar) qui a accepté de juger ce travail en tant qu'examineur. Je lui adresse mes sentiments les plus respectueux.

Un grand merci à ma famille pour son soutien constant et chaleureux pendant toute la réalisation de ce travail.

Je profite de l'occasion qui m'est donnée pour remercier l'ensemble des enseignants qui, tout au long de mon cursus universitaire, m'ont appris du mieux qu'ils pouvaient, tout ce que je sais aujourd'hui.

Enfin, j'anticipe mes remerciements à tous ceux qui vont participer de près ou de loin à la bonne réalisation de la soutenance de cette thèse.



Table des matières

| | |
|---|-----------|
| <i>Résumé</i> | 9 |
| <i>English summary</i> | 11 |
| <i>Liste des travaux</i> | 13 |
| 1 Présentation | 15 |
| 1.1 Statistique fonctionnelle | 15 |
| 1.2 Les modèles non paramétriques conditionnels comme outils d'analyse prévisionnelle . | 17 |
| 1.2.1 La régression modale | 18 |
| 1.3 Les méthodes d'estimation | 19 |
| 1.3.1 Le régressogramme | 19 |
| 1.3.2 L'estimateur de la fenêtre mobile | 20 |
| 1.3.3 L'estimateur à noyau | 20 |
| 1.3.4 L'estimation des k plus proche voisins (k -NN) | 20 |
| 1.4 Semi-métrique et petites boules de probabilités | 22 |
| 1.5 Quelques situations particulières et concrètes de données fonctionnelles | 23 |
| 1.5.1 Etude du phénomène El Niño | 23 |
| 1.5.2 Taux de graisse et courbes spectrométriques | 25 |
| 1.6 Description de la thèse | 26 |
| 1.6.1 Notations | 27 |
| 1.6.2 Présentation des résultats obtenus | 28 |
| 2 Convergence presque complète : cas <i>i.i.d</i> | 33 |
| 2.1 Introduction | 35 |
| 2.2 Models and estimators | 36 |
| 2.3 Technical tools | 38 |
| 2.4 Hypotheses and results | 40 |
| 2.5 Almost complete convergence and almost complete convergence rate | 41 |

| | | |
|----------|---|------------|
| 2.6 | <i>Real data application</i> | 42 |
| 2.7 | <i>Appendix</i> | 44 |
| 2.8 | <i>Conclusion and perspectives</i> | 52 |
| 3 | <i>Normalité asymptotique : Cas i.i.d</i> | 57 |
| 3.1 | <i>Introduction</i> | 58 |
| 3.2 | <i>The model</i> | 59 |
| 3.3 | <i>Hypotheses and results</i> | 62 |
| 3.4 | <i>Some perspectives for predictive confidence band</i> | 64 |
| 3.4.1 | <i>Prediction</i> | 64 |
| 3.4.2 | <i>Predictive confidence bands</i> | 64 |
| 3.5 | <i>Simulation study</i> | 65 |
| 3.6 | <i>Appendix</i> | 68 |
| 4 | <i>Convergence presque complète : cas α-mélange</i> | 73 |
| 4.1 | <i>Introduction</i> | 74 |
| 4.2 | <i>Notation and assumptions</i> | 77 |
| 4.3 | <i>Almost complete convergence and almost complete convergence rate</i> | 79 |
| 4.4 | <i>Real data application</i> | 82 |
| 4.5 | <i>Appendix</i> | 84 |
| 5 | <i>Normalité asymptotique : Cas α-mélange</i> | 97 |
| 5.1 | <i>Introduction</i> | 98 |
| 5.2 | <i>Notation and assumptions</i> | 102 |
| 5.3 | <i>Application</i> | 104 |
| 5.3.1 | <i>Application to prediction</i> | 104 |
| 5.3.2 | <i>Confidence bands</i> | 105 |
| 5.4 | <i>Simulation study</i> | 105 |
| 5.5 | <i>Appendix</i> | 108 |
| 6 | <i>Annexe 1 : Commentaires et perspectives</i> | 115 |
| 6.1 | <i>Quelque questions pratiques</i> | 115 |
| 6.1.1 | <i>Choix de la semi-métrique</i> | 115 |
| 6.1.2 | <i>Choix du noyau</i> | 116 |
| 6.1.3 | <i>Choix du paramètre de lissage</i> | 116 |
| 6.1.4 | <i>A propos de l'existence de la version régulière de la probabilité conditionnelle</i> | 117 |
| 6.2 | <i>Perspectives</i> | 118 |
| 6.2.1 | <i>Questions ouvertes</i> | 118 |

| | | |
|----------|--|------------|
| 7 | <i>Annexe 2 : Définitions et outils probabilistes</i> | 121 |
| 7.1 | <i>Notations et définitions</i> | 121 |
| 7.2 | <i>Outils de probabilité utilisés</i> | 122 |

Résumé

Au cours des dernières décennies, la branche de la statistique consacrée à l'étude de variables fonctionnelles a connu un réel essor tant en terme de développements théoriques que de diversification des domaines d'application. Nous nous intéresserons plus particulièrement dans cette thèse à l'estimation non paramétrique du mode conditionnel par la méthode des k -plus proches voisins dans lesquels la variable réponse est réelle tandis que la variable explicative est fonctionnelle, c'est à dire à valeurs dans un espace de dimension infinie.

L'estimation non-paramétrique par la méthode des k -plus proches voisins, appelée communément k -pp offre la particularité d'être relativement insensible aux large déviation dues à quelques observations de valeurs atypiques.

Dans ce cadre, on se propose dans cette thèse d'étudier l'estimation du mode conditionnel par la méthode des k -pp dans le cas où les observations sont de nature fonctionnelle. Plus précisément, le problème le plus connu, consiste à étudier la relation entre deux variables, afin de prédire l'une d'elles (la variable réponse) en fonction de l'autre (la variable explicative). En effet, on suppose que nous disposons d'une variable aléatoire réelle notée Y , et d'une variable fonctionnelle notée X . Le modèle non paramétrique utilisé pour étudier le lien entre X et Y concerne la distribution conditionnelle dont la fonction de densité noté f , est supposé appartenir à un espace fonctionnel approprié.

Notre étude porte sur des données indépendantes identiquement distribuées (i.i.d.) ainsi que sur des données fortement mélangeantes.

Dans un premier temps, nous considérons une suite d'observations i.i.d. Dans ce contexte, nous construisons un estimateur par la méthode des k -pp pour la fonction densité conditionnelle. Nous établissons la vitesse de convergence presque complète ainsi que la normalité asymptotique de cet estimateur.

Dans un second temps, nous supposons que les observations sont fortement mélangeantes, et nous étudierons la normalité asymptotique de cet estimateur, en donnant l'expression de sa vitesse de convergence presque complète.

Notre étude met en évidence le phénomène de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules. Cette propriété de concentration permet de contourner le problème du fléau de la dimension ainsi que la non existence d'une densité de probabilité et elle est moins restrictive que la stricte positivité de la densité de probabilité en dimension finie.

La prévision en statistique paramétrique ou non paramétrique est l'une des questions les plus cruciales auxquelles les statisticiens ne cessent de proposer des solutions dans différents contextes.

Dans le contexte non paramétrique, il est à noter que le modèle de régression usuel ne répond pas dans certaines situations aux problèmes de prévision, comme dans le cas de densité conditionnelle dissymétrique, ou, dans le cas où la densité représente plusieurs pics dont l'un d'eux est très important. Le mode conditionnel par la méthode des des k -pp est une alternative pour répondre au problème mentionné. Cette thèse s'inscrit dans la continuité des travaux existants en dimension infinie et développe aussi bien des aspects pratiques que théoriques.

Nos résultats sont appliqués à des données réelles de type spectrométrie ainsi que sur des données liées à l'étude du courant marin El Niño.

Mots clefs : donnée fonctionnelle; mode conditionnel, régression non paramétrique, l'estimateur k -pp, taux de convergence.

Summary

During the last decades, the branch of statistics dedicated to the study of functional variables has known a real development in term of theoretical developments that diversification of fields of application.

We are more particularly interested in this thesis to the non parametric estimation of the conditional mode by the k -Nearest Neighbour method, in which the response variable is real whereas the explanatory variable is functional (taking its values in an infinite dimensional space).

Nonparametric estimation by the k -Nearest Neighbour method, commonly called k -NN offer the peculiarity to be relatively insensitive to the large deviation due to some observations of atypical values.

Within this framework, one proposes in this thesis studied the conditional mode by the k -NN method where the observation are functional. More exactly, the most known problem, consists to study the relation between two variables, to predict one of them (the variable answer) according to the other one (the explanatory variable). The nonparametric model used to study the link between X and Y concern the distribution of conditional mode which the density function noted f , is supposed to belong to an appropriate functional space.

Firstly, one considers a simple of identically distributed independent observations. In this context, we build an estimator of the conditional density function, we establish the almost complete convergence rate as well as the asymptotic normality of this estimator.

Secondly, we suppose that the observations are strongly mixing, and we establish the almost complete convergence rate of the studied model. The asymptotic normality in this correlation case is also established.

Our study highlights the phenomenon of concentration properties of the probability measure of the functional variable on the small balls. This property of concentration permits to solve the problem of the curse of the dimensionality which is less restrictive to the strict positivity of the probability density in finite dimension.

The forecasting in statistical parametric or nonparametric is one of the crucial questions to which the statisticians quit to propose some solutions in different contexts.

It is not worth to say that the usual regression model doesn't answer to the problems of forecasting in some situations as in the case of asymmetric densities or in the case where the density presents

several peaks of which one of the peaks is sufficiently large. The conditional mode is alternatives to answer the mentioned problem. This thesis appears in the continuity of the existing works in infinite dimension and develop a lot of aspects of practical point of view and the theoretical one. Our results are applied to the real data of spectrometric type and to the sea current El Niño.

Key words : functional data; conditional mode, nonparametric regression, k -NN estimator, rate of convergence, random bandwidth.

Liste des travaux

Publications

1. Attouch M., Bouabça, W. (2013). The k -nearest neighbors estimation of the conditional mode for functional data. *Rev. Roumaine Math. Pures Appl.* Soumis.
2. Attouch M., Bouabça, W., Chiker El Mezoaur, Z. (2014a). The k -nearest neighbour estimation for functional nonparametric conditional mode estimator for functional time-series data. *Springer*. En révision.
3. Attouch M., Bouabça, W. (2014b). Some asymptotic normality result of k -Nearest Neighbour estimator of the conditional mode function for independent functional data. *Statistical Science*. Soumis pour publication.
4. Attouch M., Bouabça, W. (2014c). Some asymptotic normality result of k -Nearest Neighbour estimator of the conditional mode function for functional data under dependency. *Georgian Mathematical Journal*. Soumis pour publication.

Communications

1. Attouch M., Bouabça, W. The k -nearest neighbors estimation of the conditional mode for functional data. Le Congrès des Mathématiciens Algériens, Tlemcen, 11 au 13 Mai 2014.
2. Attouch M., Bouabça, W. Some asymptotic normality result of k -Nearest Neighbour estimator of the conditional mode function for functional data under dependency. Workshop on Mathematics for Life Sciences, Sidi Bel Abbès, 14 au 16 Septembre 2014.

Chapitre 1

Présentation

De presque tous les temps, et dans tous les domaines, l'homme a toujours eu besoin de comprendre les phénomènes observés et de les anticiper. C'est ce qui explique l'émergence et le succès de la statistique, dont l'objet principal est de fournir une analyse, ou une description, d'un phénomène passé, et de prédire un phénomène à venir de nature similaire. Ainsi, le travail de la statistique est d'imaginer un mécanisme probabiliste susceptible d'avoir produit les données dont elle dispose.

Concrètement, le problème le plus connu, de ceux que la statistique se propose de traiter, consiste à étudier la relation entre deux variables d'un modèle statistique, afin de prédire l'une d'elles (la variable réponse) en fonction de l'autre (la variable explicative). Il y a plusieurs façons d'étudier ces prédictions, et l'une des plus connues est certainement *la méthode de régression*, qui est basée sur l'espérance mathématique conditionnelle.

Un modèle statistique est un triplet $(\Omega, \mathcal{A}, \mathcal{P})$, où Ω est un ensemble appelé espace des observations, \mathcal{A} est une tribu sur Ω , et \mathcal{P} une famille $(P_\theta, \theta \in \Theta)$, de mesures de probabilité sur (Ω, \mathcal{A}) . Nous dirons que le modèle statistique est non paramétrique si Θ contient un ensemble convexe de dimension infinie, et qu'il est paramétrique quand Θ est un ouvert de \mathbb{R}^s avec $(0 < s < \infty)$.

Le développement considérable de la statistique non paramétrique, depuis une cinquantaine d'années, implique une intense activité dans le domaine de l'estimation fonctionnelle. Deux problèmes ont spécialement retenu l'attention des chercheurs : l'estimation des densités de probabilité, et celle des fonctions de régression.

1.1 Statistique fonctionnelle

Les données fonctionnelles sont des observations qui contiennent, au moins, un objet sous forme de fonction qui prend ses valeurs dans un espace infini-dimensionnel (appelé espace fonctionnel), comme une courbe, une surface, une image \dots etc.

Durant ces dernières années, l'utilisation des données fonctionnelles est devenue très fréquente. Grâce

au progrès technologique, les techniques infini-dimensionnelles sont maintenant plus puissantes, et la quantité des informations collectées devient volumineuse. Ces améliorations ont naturellement mené au développement de la théorie de la statistique fonctionnelle.

Il y a actuellement un nombre croissant de situations où les données collectées sont fonctionnelles. Des situations provenant de différents domaines des sciences appliquées, aussi bien environnementales que chimiques, biométriques, médicales, ou même économiques. En effet, le progrès des outils informatiques, que ce soit en terme de capacités mémorielles ou informationnelles, nous a permis de traiter de très grands ensembles de données.

Les courbes de croissance, les enregistrements sonores, les images satellites, les séries chronologiques, les courbes spectrométriques ne sont que quelques exemples illustrant le grand nombre et la diversité des données de nature fonctionnelle auxquelles le statisticien peut être confronté. C'est une des raisons pour lesquelles un nouveau champ de la statistique, dédié à l'étude de données fonctionnelles, a suscité un fort engouement au début des années quatre-vingt, sous l'impulsion, notamment, des travaux de Grenander (1981), Dauxois et al. (1982) et Ramsay (1982). Il a été popularisé par Ramsay et Silverman (1997), puis par les différents ouvrages de Bosq (2000), Ramsay et Silverman (2002, 2005), et Ferraty et Vieu (2006a). C'est un domaine de la Statistique en plein essor comme en témoignent les hors-séries qui lui sont ou seront consacrés dans des revues reconnues comme *Statistica Sinica* (2004), *Computational Statistics & Data Analysis* (2006), *Computational Statistics* (2007), *Journal of Multivariate Analysis* (2008). De plus, même si les données dont dispose le statisticien ne sont pas de nature fonctionnelle, il peut être amené à étudier des variables fonctionnelles construites à partir de son échantillon initial. Un exemple classique est celui où l'on observe plusieurs échantillons de données réelles indépendantes et où l'on est ensuite amené à comparer les densités de ces différents échantillons ou bien à considérer des modèles où elles interviennent.

Récemment, Gasser et al (1998), Hall et Heckman (2002) se sont intéressés à l'estimation non paramétrique du mode de la distribution d'une variable fonctionnelle, Cadre (2001) a étudié l'estimation de la médiane qui prend ses valeurs dans un espace de Banach.

Ferraty et al (2004) ont étudié l'estimateur non paramétrique du mode de la densité d'une variable aléatoire à valeurs dans un espace vectoriel semi-normé de dimension infinie. Ils établissent sa convergence presque sûre et ils appliquent également ce résultat au cas où la mesure de probabilité de la variable vérifie une condition de concentration. Plusieurs auteurs se sont intéressés à l'application de la modélisation statistique par des variables fonctionnelles sur des données réelles voir Ferraty et Vieu (2002, 2003) en données spectrométriques, Besse et al. (2002) en météorologiques, Gasser et al. (1998) en médecine.

Le premier résultat sur la normalité asymptotique sur l'estimateur à noyau vient de Masry (2005), Il considère le cas de données α -mélangeant, mais il n'a pas donné l'expression explicite des termes dominants asymptotiques de biais et de variance. Après Ferraty et al. (2007) donnent l'expression explicite de la loi asymptotique (c'est-à-dire des termes dominants de biais et de variance) dans le cas d'un ensemble de données indépendant. Les résultats exposés dans les papiers Delsol

(2007a, 2008b) font le lien entre ces articles, ils généralisent les résultats de Ferraty et al. (2007) au cas de données α -mélange, Attouch et Benchikh (2012) ont établi la normalité asymptotique de l'estimation robuste de la fonction de régression.

1.2 Les modèles non paramétriques conditionnels comme outils d'analyse prévisionnelle

Une question très importante en statistique c'est d'étudier les liens entre deux variables aléatoires. D'un point de vue historique, ce problème a été abordé pour la première fois dans un contexte géométrique par Galileo Galilei en (1632). L'idée principale de ce dernier est d'ajuster un nuage de point par une droite permettant d'interpréter la relation entre des données contaminées. Une formulation mathématique pour ce problème, connue dans la littérature sous le nom de régression linéaire.

Le principe de la régression non paramétrique a vu le jour au 19^{me} siècle. Toutefois, les premiers travaux modernes sur ce sujet datent des années 50. Ces premiers travaux concernaient l'estimation de la fonction de densité par des méthodes d'opérateurs à noyaux, avec les travaux fondateurs de Rosenblatt (1956), Tukey (1961) et Parzen (1962). Ils ont ensuite été étendus à l'estimation de la régression par opérateur à noyau, travaux publiés pour la première fois en (1964), indépendamment, par Nadaraya et Watson. En (1988), Cleveland et Develin ont généralisés ces méthodes par la régression locale polynomiale.

En effet, la régression non paramétrique est un outil statistique permettant de déterminer la variation de l'espérance mathématique d'une variable dépendante Y en fonction d'une ou de plusieurs variables explicatives, sans spécifier une forme stricte pour cette relation. En d'autres termes, nous étudierons comment Y évolue "en moyenne" en fonction de ces variables explicatives sans avoir de fonction stricte pour décrire cette évolution.

Ainsi, si nous disposons de n couples indépendants de variables aléatoires $(X_i, Y_i)_{i=1, n}$ tels que :

$$Y_i = r(X_i) + \epsilon_i, \tag{1.1}$$

où :

- r une fonction inconnue, mesurable par rapport à l'observation $(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n)$, et non nécessairement linéaire. Cette fonction est appelée fonction de régression.
- ϵ_i est le résidu ; il exprime tous les facteurs explicatifs autres que les variables X_i , que nous ne savons pas, ou que nous ne voulons pas, prendre en compte. Nous dirons que les ϵ_i sont inobservables, et nous supposerons qu'ils sont indépendants des X_i et que : $\epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad \forall i = 1, \dots, n$.

Nous pouvons aussi exprimer la formule (1.1) comme suit :

$$r(x) = \mathbb{E}(Y_i/X_i = x). \quad (1.2)$$

Le problème de la régression non paramétrique est celui de l'estimation de la fonction de régression r sachant qu'elle appartient à un ensemble de dimension infini, et que nous ne pouvons donc l'écrire ni sous une forme explicite, ni en fonction d'un nombre réduit de paramètres.

Les paramètres fonctionnels conditionnels, tels la distribution conditionnelle, la densité conditionnelle, le mode conditionnel, le quantile conditionnel, la fonction de hasard conditionnelle, sont largement étudiés en dimension finie. Via ces paramètres, la prévision qui est une question très fréquente en statistique dans les modèles non paramétriques offre une véritable alternative à la régression non paramétrique où cet outil a pris un essor considérable de par le nombre de publication qui lui sont consacrées, que les variables explicatives soient uni, multi ou infiniment dimensionnelles, Cependant cet outil de prévision n'est pas très adapté pour certaine situation.

En dimension finie, il existe une littérature abondante pour ces paramètres conditionnels. Roussas (1968) fut le premier à établir des propriétés asymptotiques pour l'estimateur à noyau de la fonction de distribution conditionnelle, pour des données Markoviennes, il montre ainsi la convergence en probabilité. Youndjé (1993) s'est intéressé à l'étude de la densité conditionnelle pour des données dépendantes et indépendantes. Laksaci et Yousfate (2002) ont établi, pour un processus markovien stationnaire, la convergence en norme L^p pour l'estimateur à noyau de la densité conditionnelle.

Dans la suite de ce paragraphe, nous définissons cet outil, et en mettant l'accent sur la diversité et l'étendue de la littérature disponible sur ce modèle.

1.2.1 La régression modale

Un autre prédicateur alternatif à la régression classique est le mode conditionnel. Par définition, le mode conditionnel est la valeur qui maximise la densité conditionnelle. Nous supposons qu'il existe un compact \mathcal{S} ou le mode est unique note θ . Ce prédicateur est obtenu, en considérant :

$$\theta(x) = \arg \max_{y \in \mathbb{R}} f^x(y). \quad (1.3)$$

où $f^x(y)$ est la densité conditionnelle de y sachant x .

L'estimation du mode ainsi que le mode conditionnel, et vu l'intérêt de ces paramètres dans le domaine de la prévision, plusieurs auteurs s'y sont intéressés, citons, par exemple, Parzen (1962) qui établit pour des données réelles i.i.d., la normalité asymptotique sous de faibles conditions, le même résultat fut établi par Eddy (1980) sous des conditions fortes. Dans le cas multidimensionnel, les mêmes résultats ont été obtenus par Samanta (1973) et Konakov (1974). Romano (1988) a également étudié le comportement asymptotique du mode sous des hypothèses de régularité de la fonction de densité. Vieu (1996) obtient une vitesse de convergence pour l'estimateur local et global de la fonction mode. Enfin, Louani (1998) établi la normalité asymptotique pour la densité et ses dérivées

avec application au mode.

Pour le mode conditionnel, Samanta et Thavaneswaran (1990) ont établi, pour des données i.i.d. sous certaines conditions de régularités la normalité asymptotique. Dans le cas des données dépendantes, Collomb et al. (1987) ont fourni des propriétés asymptotiques sous condition de α -mélange .

Ces résultats ont été étendus respectivement aux cas α -mélange et ergodique par Ould-Saïd (1993, 1997)), Quintela et Vieu (1997) ont estimé le mode conditionnel comme étant le point annulant la dérivée d'ordre un de l'estimateur de la densité conditionnelle, et ont établi la convergence presque complète de cet estimateur sous la condition α -mélange. Louani et Ould-Saïd (1999) ont établi la normalité asymptotique dans le cas des données fortement mélangeantes. Dans le cas des données censurées, Ould-Saïd et Cai (2005) ont établi la convergence uniforme sur un compact pour l'estimateur à noyau du mode conditionnel.

En dimension infinie, le mode conditionnel connaît tout récemment, un intérêt croissant, malgré le peu de résultats existants. Dans le contexte des données fonctionnelles, les premiers travaux ont été obtenus par Ferraty, Laksaci et Vieu (2006). Ils ont montré, sous des conditions de régularités de la densité conditionnelle, la convergence presque complète de l'estimateur à noyau qu'ils ont introduit pour la densité conditionnelle ainsi que pour le mode conditionnel, et ont établi la vitesse de convergence. Une application de leurs résultats aux données issues de l'industrie agro-alimentaire a été présentée.

Dans le même contexte et concernant le mode, Dabo-Niang, Ferraty et Vieu (2006) ont étudié l'estimateur à noyau de mode, pour une variable fonctionnelle à valeur dans un espace normé de dimension infinie. Ils ont établi la convergence uniforme presque complète, et ont spécifié les vitesses de convergence en montrant le poids de la mesure des petites boules sur ces dernières. Une application aux données spectrométriques a été présentée.

1.3 Les méthodes d'estimation

Faisons une brève présentation de l'évolution des méthodes de l'estimation non paramétrique, avant d'introduire la méthode qui fait l'objet de cette thèse.

1.3.1 Le régressogramme

En (1961), Tukey a défini un estimateur pour la régression, comme suit : considérons une partition $(C_j)_{j=1,\dots,J}$ du support de X , fixée à priori. Pour la classe C_j où se trouve le point x , on effectue la moyenne des Y_i correspondants aux X_i de cette même classe C_j . Autrement dit :

$$\forall x \in C_j : \hat{r}_{reg}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{X_i \in C_j\}}}{\sum_{i=1}^n \mathbb{I}_{\{X_i \in C_j\}}} .$$

Mais cet estimateur primitif présente comme inconvénients le fait d'avoir à choisir, à la fois, la finesse de la discrétisation (i.e. le nombre J de découpages) et la position exacte des bornes des intervalles C_j .

1.3.2 L'estimateur de la fenêtre mobile

Afin de remédier à ce problème, un nouvel estimateur a été construit en remplaçant la partition en intervalles C_j par un seul intervalle, mais qui varie de manière continue. Nous obtenons ainsi :

$$\forall x : \hat{r}_{FM}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{X_i \in [x-h; x+h]\}}}{\sum_{i=1}^n \mathbb{I}_{\{X_i \in [x-h; x+h]\}}}, \quad \text{avec } h \in \mathbb{R}_+^*.$$

Cet estimateur, aussi, présente des désavantages, vu qu'il est constant sur chaque intervalle $[x-h; x+h]$, et donc discontinu.

1.3.3 L'estimateur à noyau

C'est ainsi qu'en (1964), Nadaraya et Watson ont eu l'idée (indépendamment et simultanément) de définir un estimateur à noyau comme suit :

$$\forall x : \hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)},$$

où K est une "fonction noyau", et h un réel strictement positif appelé "paramètre de lissage".

1.3.4 L'estimation des k plus proche voisins (k -NN)

L'estimation des k plus proche voisins (ppv) était largement étudiée si la variable explicative est un élément d'un espace fini-dimensionnel, voir Györfi et al (2002). Dans le cas fonctionnel avec une réponse réelle-estimé, deux approches différentes pour l'estimation de la fonction de régression par k -NN existent. Le premier travail, a était publié par Laloë (2008), et a traité une estimation non paramétrique par k -NN quand la variable fonctionnelle est un élément d'un espace Hilbert séparable. Pour ce cas Laloë a établi un résultat de convergence faible. Cependant, son ansatz n'est pas complètement fonctionnel. La stratégie de Laloë est de réduire la dimension de la variable d'entré en utilisant une projection sur un sous-espace de dimension fini et appliquant ensuite une technique multi variée sur les données projetées. Le deuxième résultat, de Burba et al (2009), est ayant été sur une approche fonctionnelle pure. Ils ont examiné le problème sur un espace fonctionnel semi-métrique. La convergence et le taux de convergence presque complets sont obtenus pour des

données indépendantes. En outre, Burba et al ont prolongés un Lemme de Collomb (1980) que nous utiliserons ainsi dans nos démonstrations. De plus, l'estimation à noyau par les k -NN est examinée pour la classification dans la dimension infinie par Cérout et Guyader (2003) et il existe un résultat de convergence pour l'estimation de régression par les k -NN quand la réponse est un élément d'un espace de Hilbert (voir Lian (2011)).

Dans le cas d'une variable explicative de dimension finie, l'estimation à noyau des k -NN pour des variables aléatoires α -mélangeant est traité par Tran (1993) et Lu et Cheng (1998).

L'idée principale de cette méthode est basée sur l'estimation d'un nombre fixe d'observations k qui est le plus proche au point désiré, supposons $X \in R^q$ et nous avons un échantillon $\{X_1; \dots; X_n\}$.

Pour n'importe quel point fixe $x \in R^q$, nous pouvons calculer à quelle distance chaque observation X_i est à x on utilisant la distance euclidienne $\|x\| = (x'x)^{\frac{1}{2}}$. Cette distance est

$$D_i = \|x - X_i\| = ((x - X_i)'(x - X_i))^{\frac{1}{2}},$$

ceci est juste un calcul simple sur l'ensemble de données.

La statistique d'ordre pour les distances D_i sont $0 \leq D(1) \leq D(2) \leq \dots \leq D(n)$.

Les observations correspondantes à cette statistique d'ordre sont les "plus proches voisins" de x . Le premier plus proche voisin est l'observation la plus proche de x , le deuxième plus proche voisin est la seconde observation la plus proche, etc.

Les observations classées par les distances, où les plus proches voisins sont $\{X_{(1)}; X_{(2)}; X_{(3)}; \dots; X_{(n)}\}$, le k -ème plus proche voisin de x est $X_{(k)}$.

Pour une donné k ; nous définissons

$$R_x = \|X_{(k)} - x\| = D_{(k)},$$

R_x est juste le k -ème ordre statistique sur les distances D_i .

Notre recherche développée dans cette thèse est basée sur les premiers travaux qui ont été réalisés par Royall en (1966) et Stone en (1977) : l'idée est de ne considérer que les Y_i correspondants aux k points X_i les plus proches de x ($k \in \mathbb{N}^*$), pour l'estimation de la régression.

Pour ce faire, nous remplaçons le paramètre de lissage h de la méthode des noyaux, par une variable aléatoire positive $H_{n,k}$ qui représente la k -ème plus grande valeur des distances $|x - X_i|$.

Le premier avantage de la méthode k -NN vient de la nature de son paramètre de lissage k qui prend ses valeurs dans un ensemble discret $F = \{1, 2, 3, \dots, n\}$, tandis que, dans la méthode des noyaux, le paramètre de lissage h est un nombre réel positif.

Le second avantage, et le plus important, réside dans le fait que cette méthode permet la construction d'un voisinage adapté à la structure locale des données. En effet, quand les données sont hétérogènes, c'est à dire de natures différentes, la méthode des noyaux classique n'est pas très efficace pour estimer la fonction de régression. Alors que dans la méthode k -NN, le fait de choisir le nombre des voisins permet d'éviter ce problème.

La méthode k -NN est encore plus efficace, quand les données sont fonctionnelles, grâce au paramètre

adaptatif k .

Toutefois, la méthode k -NN présente une difficulté technique majeure. À savoir que la sélection des plus proches voisins donne une variable aléatoire. En effet, ce caractère aléatoire crée des difficultés techniques pour les démonstrations, puisque nous ne pouvons pas utiliser les mêmes outils que dans la méthode des noyaux standard.

1.4 Semi-métrique et petites boules de probabilités

La problématique de dimension est un phénomène célèbre dans la régression non paramétrique sur la variable multi variée (voir Stone(1982)). Dans la régression non-paramétrique multi variée, les taux de convergence (pour la partie de dispersion) sont exprimé en termes de h_n^d . Dans le cas fonctionnel nous adoptons des notions de concentration plus générales appelées petites boule de probabilités et exprimons nos résultats asymptotiques en fonction de ces quantités. Les petites boule de probabilités sont définies par :

$$F_x(h_n) = \mathbb{P}(d(X, x) \leq h_n).$$

La façon de diminuer au zéro à une grande influence sur le taux de convergence de l'estimation par la méthode des k -NN . Nous pouvons trouver dans beaucoup d'articles de probabilité des équivalents asymptotiques pour ceux des petites boule de probabilités quand d est une norme (voir par exemple Bogachev (1998)), Li et Shao (2001), Dereich (2003), Lifshits et al (2006), Shmileva(2006), Gao et Li(2007) ou une semi-norme spécifique Lifshits et Simon(2005), Aurzada et Simon(2007). Dans le cas de processus non lisses (mouvement Brownian, processus Ornstein- Ullenberg...), ces petites boules de probabilités ont une forme exponentielle (en ce qui concerne h_n) et par conséquent le taux de convergence est une puissance de $\log(n)$ (voir Ferraty et Vieu(2006a)). Le choix de la semi-métrique a une influence directe sur la topologie et par conséquent sur la petite boule de probabilités. La diversité de semi métrique permet, dans des situations diverses, de trouver une topologie qui donne une notion pertinente de proximité entre les courbes (voir Ferraty et Vieu(2006a)). Par exemple, le choix d'une projection semi-métrique (basé sur des composants principaux fonctionnels, la décomposition de Fourier, splines...) constitue une alternative intéressante à la problématique de dimension. Quand E est un espace Hilbert séparable, Ferraty et Vieu(2006a)(Lemme 13.6, p.213), montrent qu'il existe un spécifique semi-métrique qui mène aux petites boule fractal de probabilités. (*i.e.* $\exists C, \delta > 0, \phi_x(h)h \stackrel{h \rightarrow 0}{\sim} C_x h^\delta$) où le tau de convergence de l'estimateur à noyau est une puissance de n . Dans d'autres situations, typiquement identique, quand l'ensemble de données sont composées de courbes lisse (comme les courbes spectrométriques) cela peut être intéressant d'utiliser la semi métrique basée sur des dérivées.

Nous pouvons se demander comment choisir la semi-métrique en pratique. Ferraty et al (2002b) ont proposés la première méthode, premièrement on doit choisir une famille de semi-métrique du de l'information arrive sur les données, alors, par exemple on détermine, par la cross-validation, la

semi-métrique (parmi cette famille) qui est le plus adapté aux données. La justification théorique de l'utilité d'un semi-métrique particulier est toujours un problème ouvert.

1.5 Quelques situations particulières et concrètes de données fonctionnelles

Provenant de domaines assez variés (climatologie et chimométrie, électrique), nous présentons deux jeux de données fonctionnelles particuliers par une approche fonctionnelle pour illustrer quels sont les avantages et les enjeux liés à leur étude.

1.5.1 Etude du phénomène El Niño

El Niño est le phénomène climatique le plus spectaculaire : il se développe dans le Pacifique Equatorial vers Noël (d'où son nom, qui signifie Enfant Jésus en espagnol), mais on sait depuis une bonne dizaine d'années que son impact sur le climat est global et touche non seulement toute la ceinture équatoriale, mais aussi les latitudes tempérées, en particulier le climat en Amérique du Nord et dans les régions asiatiques, en relation avec la mousson.

Les mécanismes qui régulent ce phénomène ne sont pas encore entièrement connus, mais ils font intervenir les interactions couplées entre l'atmosphère et l'océan : surtout dans les régions équatoriales, où la température de l'océan est très élevée, l'atmosphère et l'océan sont en continuelle interaction, la température de surface affectant les vents, qui eux même affectent les courants et la température de surface. El Niño apparaît environ une à deux fois par décennies, avec une intensité particulièrement importante lors de ces dernières manifestations et sa durée est d'environ 18 mois. Le jeu de données dont nous disposons est constitué de relevés de température mensuels de la surface océanique effectués depuis 1950 dans une zone située au large du nord du Pérou (aux coordonnées $0 - 10^\circ$ Sud, $80 - 90^\circ$ Ouest) où peut apparaître le courant marin El Niño.

Ces données et leur description sont disponibles sur le site du centre de prévision du climat américain : <http://www.cpc.ncep.noaa.gov/data/indices/>.

L'évolution des températures au cours du temps est réellement un phénomène continu. Le nombre de mesures dont nous disposons en rend assez bien compte (voir Figure 1.1) et permet de prendre en considération la nature fonctionnelle des données.

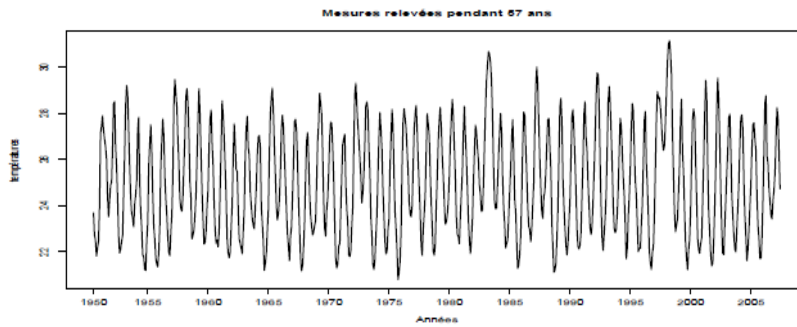


FIGURE 1.1 – Mesures mensuelles de la température à la surface autour du courant marin El Niño depuis 1950.

À partir de ces données on peut s'intéresser à essayer de prédire l'évolution du phénomène à partir des données recueillies lors des années précédentes.

Plusieurs approches ont été introduites pour tenter de répondre à ce problème. Les premiers travaux cherchent à prédire la température du mois suivant à partir des p températures mensuelles précédentes (pour plus de références voir Katz, 2002).

Toutefois cette modélisation ne permet pas de prendre en compte la nature fonctionnelle du phénomène étudié ni d'en tirer profit. Plus récemment, suite à l'approche introduite par Bosq (1991), on a choisi de considérer le processus non plus au travers de sa version discrétisée mais bien comme un processus continu que l'on décompose en courbes successives de 12 mois (entre juin d'une année et mai de l'année suivante, (voir Figure1.2)).

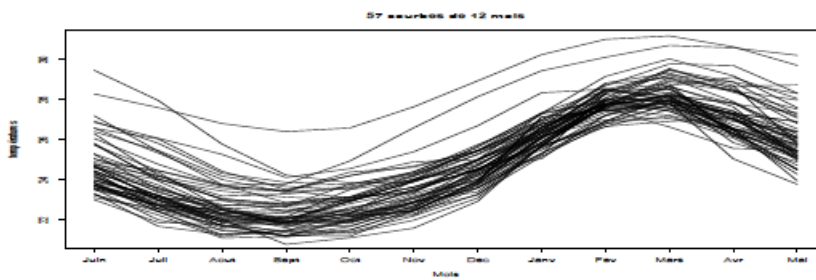


FIGURE 1.2 – 57 courbes représentant la température à la surface autour du courant marin El Niño par tranches de 12 mois depuis juin 1950.

À partir d'une ou plusieurs courbes précédentes, on a cherché à prédire la courbe des douze mois suivants (ou certaines valeurs particulières de celle-ci). Cette approche fonctionnelle a l'avantage de

rester adaptée même lorsque le nombre de points de discrétisation augmente et permet de prendre en compte, pour chaque période de douze mois, la courbe des températures dans sa globalité. Les travaux de Besse et al. (2000), Valderama et al. (2002), Antoniadis et Sapatinas (2003) ou Ferraty et al. (2005) illustrent différentes manières de répondre à ce problème et de modéliser la dépendance des courbes successives. Les résultats donnés dans la première partie de ce mémoire sont établis pour des variables fonctionnelles faiblement dépendantes et peuvent eux aussi être appliqués pour répondre à des problématiques de prédiction de séries temporelles.

1.5.2 Taux de graisse et courbes spectrométriques

L'abondance des jeux de données fonctionnelles actuellement disponibles offre un grand choix de possibilités d'applications au statisticien désireux d'illustrer le comportement d'une nouvelle méthode. Étant donné que notre propos est d'illustrer la complémentarité des diverses approches proposées dans la littérature il est naturel de concentrer notre travail sur l'exemple qui a été le plus souvent étudié. Cet exemple concerne l'industrie agro-alimentaire et plus particulièrement un problème de contrôle de qualité sur de la viande hachée. La variable fonctionnelle est donnée par la courbe d'absorbance de la lumière en fonction de la longueur d'onde, courbe obtenue au moyen d'une technique classique (et peu onéreuse) de spectrométrie dans le proche infrarouge. Le jeu de données dont nous disposons est constitué de 215 courbes d'absorbance (voir Figure 1.3).

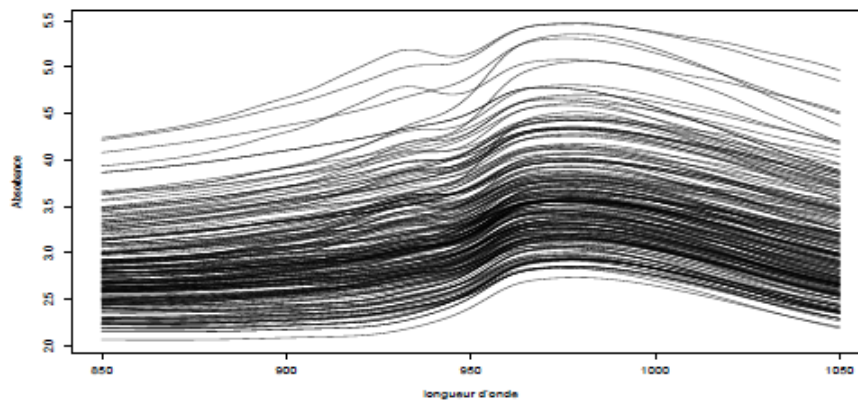


FIGURE 1.3 – Courbes spectrométriques obtenues à partir des 215 morceaux de viande étudiés.

Ces données ont été abondamment étudiées dans la littérature depuis l'article de C. Borggaard et H.H. Thodberg (1992) et nous renvoyons à l'article de F. Ferraty and P. Vieu.(2006) pour une présentation plus détaillée. En présence d'un tel échantillon, les problèmes statistiques qui peuvent se poser sont les mêmes qu'en statistique multi variée classique. Il peut s'agir de problème de type

descriptif du type peut-on classer ces courbes (et comment) en plusieurs catégories ?, et dans ce cas des techniques de classification devront être développées. Il peut s'agir de problèmes de prédiction, Peut-on prédire à partir de la courbe \mathcal{X} une autre caractéristique (fonctionnelle ou non) Y ? ; et dans ce cas des techniques de régression (resp. de discrimination) devront être développées si Y est quantitative (resp. si Y est qualitative). Face à d'autres questions du type certaines parties de la variable \mathcal{X} , par exemple certaines régions du spectre ou simplement quelques points de discrétisations bien choisis, seraient-ils plus informatifs que d'autres ?. Les techniques statistiques à développer pourront s'inspirer des méthodes de sélection de variables.

1.6 Description de la thèse

Le large champ d'applications qu'offre l'estimation fonctionnelle tant par la capacité de modéliser ces données comme des réalisations d'une variable aléatoire dans un espace abstrait que par l'outil théorique qu'elle offre, d'où l'engouement de nombreux auteurs dans différents domaines, en particulier dans le domaine de la prévision.

Le travail développé dans cette thèse se situe à l'intersection entre les deux thématique importantes : variables fonctionnelles et prévision non paramétrique via le mode conditionnel par la méthode des k -NN.

Cette thèse est divisée en sept chapitres. Le premier est consacré à la présentation des domaines d'applications des données de type fonctionnelles. Ainsi, nous avons cité quelques exemples d'application tels que la discrimination des courbes.

Le second chapitre s'intéresse à une suite d'observations i.i.d. Dans ce contexte, nous construisons un estimateur par la méthode des k -NN pour la fonction densité conditionnelle où la vitesse et le taux de convergence presque complète seront établies. Une application sur des données spectrométrique illustre la qualité de notre estimation.

Le troisième chapitre traite le problème de la normalité asymptotique pour des variables i.i.d. Une application sur des données réelles exhibe l'efficacité de notre étude.

Le quatrième chapitre traite la convergence presque complète de notre estimateur au cas de variables aléatoires dépendantes. Une application sur des données du courant marin El Niño sera étudié.

Dans le cinquième chapitre, on étend les résultats du chapitre trois au cas de variables aléatoires alpha mélangantes. Nous établissons la normalité asymptotique. On trouvera aussi dans ce chapitre un exemple sur des données réelles dont l'objectif est l'étude comparative entre le mode conditionnel k -NN et le mode conditionnel classique.

Dans le sixième chapitre, des commentaires sur le rôle des paramètres qui interfèrent dans la qualité de l'estimation sont mis en valeurs. Un septième chapitre aussi, regroupe un ensemble d'outils probabilistes utilisés dans cette thèse.

1.6.1 Notations

Soient $(X_i, Y_i)_{i=1, \dots, n}$, n couples aléatoires indépendants et de même loi que le couple (X, Y) , pris dans $F \times \mathbb{R}$.

On note d la semi-métrique de F . Pour tout $x \in F$, on définit la fonction de répartition conditionnelle de Y sachant $X=x$, noté F^x par

$$\forall y \in \mathbb{R} \quad F_Y^X = \mathbb{P}(Y \leq y | X = x).$$

On suppose que cette distribution conditionnelle de Y sachant X est absolument continue par rapport à la mesure de Lebesgue dont on note f^x sa densité. On désigne par $f^{x(j)}$ la dérivée d'ordre j de cette densité conditionnelle, comme $f^x = f^{x(0)}$, tous nos résultats seront énoncés en fonction de j . Pour contourner ce problème, introduisons la notion :

$$\phi_x(h) = P(X \in B(x, h)) > 0$$

où ϕ_x représente la mesure de concentration.

On construit un estimateur pour la fonction de répartition par la méthode des (k -NN) noté \widehat{F}_{kNN}^x

$$\widehat{F}_{kNN}^x(y) = \frac{\sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))G(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))}.$$

La fonction densité f on x par les (k -NN), où $\widehat{f}_{kNN}^{x(j)}$ est défini par :

$$\widehat{f}_{kNN}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))G^{(j+1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))}$$

où :

- K est un noyau.
- et $H_{n,k}$ est définie comme suit :

$$H_{n,k} = \min \left\{ h \in \mathbb{R}^+ / \sum_{i=1}^n \mathbb{I}_{\{\|x - X_i\| < h\}}(X_i) = k \right\},$$

ce qui définit une variable aléatoire réelle positive représentant la k -ème plus grande valeur des $\|x - X_i\|$.

- G est un noyau .
- h_G est une suite de réels positifs.

On étudie le mode conditionnel. Un estimateur de θ est la variable aléatoire $\widehat{\theta}_{kNN}$ solution du problème

d'optimisation suivant :

$$\widehat{f}^x(\widehat{\theta}_{kNN}) = \sup_{y \in \mathcal{S}} \widehat{f}_{kNN}^x(y) \quad (1.4)$$

Pour tout x fixé dans \mathcal{F} , on note par N_x , un voisinage de ce point, \mathcal{S} un compact de \mathbb{R} .

Considérons deux sortes de modèles non-paramétriques :

i) Modèle du type continu ; définit comme :

$$f \in C_E^0 = \{f : E \rightarrow \mathbb{R} / \lim_{\|x-x'\| \rightarrow 0} f(x) = f(x')\},$$

ii) modèle du type Hölder ou Lipschitzien ; définit par : $\exists \alpha > 0$ tel que :

$$f \in Lip_{E\alpha} = \{f : E \rightarrow \mathbb{R} / \exists C > 0, \forall x' \in E : |f(x) - f(x')| < C d(x, x')^\alpha\}.$$

1.6.2 Présentation des résultats obtenus

Résultat : Cas i.i.d

Nous supposons pour commencer que l'échantillon que nous étudions est constitué de variables indépendantes et nous établissons la convergence presque complète (p.co) de notre estimateur, en donnant l'expression explicite des termes de vitesse de convergence.

Théorème 1.6.1

Sous des hypothèses de concentration de la mesure de probabilité de la variable fonctionnelle ainsi que d'autres hypothèses de régularités et techniques, on a

$$\lim_{n \rightarrow \infty} \widehat{\theta}_{kNN}(x) = \theta(x) \quad p.co.$$

$$\widehat{\theta}_{kNN}(x) - \theta(x) = O\left(\phi_x^{-1} \left(\frac{k}{n}\right)^{b_1} + h_G^{b_2}\right)^{\frac{1}{j}} + O\left(\frac{\log(n)}{h_G k}\right)^{\frac{1}{2j}} \quad p.co.$$

Ce résultat sera établi dans le chapitre 2 de cette thèse.

Théorème 1.6.2

$$\left(\frac{h_G^3 k (a_1 \widehat{f}^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))}\right)^{1/2} \left(\widehat{\theta}_{kNN}(x) - \theta(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ quand } n \rightarrow \infty.$$

$\xrightarrow{\mathcal{D}}$ veut dire la convergence en distribution, $\Upsilon = \{x : x \in \mathcal{S}, f^{x(1)}(\theta(x)) \neq 0\}$ et

$$\sigma^2(x, \theta(x)) = a_2 f^{x(1)}(\theta(x)) \int_{\mathbb{R}} \left(G^{(2)}(t)\right)^2 dt \quad (\text{avec } a_j = K^j(1) - \int_0^1 (K^j)'(s) \zeta_0(s) ds, \text{ pour } j = 1, 2).$$

La démonstration et les détails des conditions imposées pour aboutir à ces résultats sont donnés dans le chapitre 3.

Résultat : Cas α mélangeant

Pour l'extension de nos résultats à des problèmes non standard de la statistique non paramétrique, en renforçant les hypothèses de concentration sur la loi conjointe (X_i, Y_i) ainsi que les hypothèses sur les coefficients de mélange, on montre le résultat suivant :

Théorème 1.6.3

Sous des hypothèses de concentration de la mesure de probabilité de la variable fonctionnelle ainsi que d'autres hypothèses de régularités et techniques, on a

$$\lim_{n \rightarrow \infty} \widehat{\theta}_{kNN}(x) = \theta(x) \quad p.co.$$

$$\widehat{\theta}_{kNN}(x) - \theta(x) = O\left(\phi_x^{-1} \left(\frac{k}{n}\right)^{b_1} + h_G^{b_2}\right)^{\frac{1}{j}} + O\left(\frac{\log(n)}{h_G k}\right)^{\frac{1}{2j}} + O\left(\frac{n^{1+s} \log(n)}{h_G k^2} \mathcal{X}\left(x, \phi_x^{-1} \left(\frac{k}{n}\right)\right)^{1-s}\right)^{\frac{1}{2j}} \quad p.co.$$

où $\mathcal{X}(x, h_K)$ est le maximum de concentration entre la loi marginale et les lois conjointes de chaque couple d'observations fonctionnelles dans la boule de centre x et de rayon h_K .

Les démonstrations et les conditions imposées afin d'obtenir ce théorème sont données au chapitre 4.

Théorème 1.6.4

$$\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left(\widehat{\theta}_{kNN}(x) - \theta(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ quand } n \rightarrow \infty.$$

$\xrightarrow{\mathcal{D}}$ veut dire la convergence en distribution, $\Upsilon = \{x : x \in \mathcal{S}, f^x(\theta(x)) \neq 0\}$ et

$$\sigma^2(x, \theta(x)) = \frac{a_2(x) f^x(\theta(x))}{[a_1(x) f^{x(2)}(\theta(x))]^2} \int_{\mathbb{R}} \left(G^{(2)}(t)\right)^2 dt.$$

La démonstration et le détail des conditions imposées font l'objet d'un article récemment soumis et qui est présenté dans le chapitre 5 de cette thèse.

Références

1. Attouch, M.K, and Benchikh, T.(2012). Asymptotic distribution of robust k-nearest neighbour estimator for functional nonparametric models. *Mathematic Vesnic*, **64**, no. 4, 275-285.
2. Aurzada, F., and Simon, T.(2007). Small ball probabilities for stable convolutions. *ESAIM Probab. Stat*, **11** 327-343 (electronic).
3. Bogachev, V.I. (1998). Gaussian measures. *Mathematical Surveys and Monographs 62 American Mathematical Society, Providence, RI*. xii+433 pp.
4. Bosq, D. (2000). Linear processes in function spaces. *Theory and Application, Lectures Notes in Statistics. Springer Verlag, New-York*, Vol. **149**.
5. Burba, F., Ferraty, F. and Vieu, P.(2008). Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique. *C. R. Math. Acad. Sci. Paris*, **346**, no. 5-6, 339-342.
6. Burba, F., Ferraty, F. and Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression. *J. Nonparametric. Stat*, **21**, no. 4, 453-469.
7. Cérou, F., and Guyader, A. (2005). Nearest neighbor classification in infinite dimension. *ESAIM : P&S* **10**, 340-255.
8. Collomb, G. (1980). Estimation de la régression par la méthode des k points les plus proches avec noyau : quelques propriétés de convergence ponctuelle, in Nonparametric Asymptotic Statistics. *Lecture notes in mathematics, Springer-Verlag*, **821**, 159-175.
9. Collomb, G., Härdle, W., and Hassani, S.(1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf*, **15**, 227-236.
10. Dabo-Niang, S., and A. Laksaci,(2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C. R., Math., Acad. Sci. Paris*, **344**, 49-52.
11. Dereich, S. (2002). Estimation de la densité dans un espace de dimension infinie : Application aux diffusions. *C. R. Acad. Sci., Paris*, **334**, No.3, 213-216.
12. Dereich, S. (2003). High resolution coding of stochastic processes and small ball probabilities. PhD Thesis.
13. Dereich, S. (2004). Kernel density estimator in an infinite dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett*, **17** :381- 386.
14. Dereich, S. , Ferraty, F., and Vieu, P.(2006). Mode estimation for functional random variable and its application for curves classification. *Far East J. Theoretical Statist*, **18**, 93-119.
15. Devroye, L.P. (1978). The uniform convergence of nearest neighbour regression function estimators and their application in optimization. *IEEE Trans. Inform. Theory*, **24**, 142-151.
16. Eddy, W. F. (1982). The asymptotic distribution of kernel estimators of the mode. *Z. W. Giebete*, **59**, 279-290.

17. Ezzahrioui, M., and Ould-Saïd, E. (2006). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. Preprint, LMPA No 277, Univ. du Littoral Côte d'Opale.
18. Ezzahrioui, M., and Ould-Saïd, E. (2008). Asymptotic normality of nonparametric estimator of the conditional mode for functional data. *Journal of Nonparametric Statistics*, **20**, 3-18.
19. Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist. and Data Anal*, **4** :545-564.
20. Ferraty, F. and Vieu, P. (2006a). Nonparametric modelling for functional data. *Springer-Verlag, New York*.
21. Ferraty, F. et Vieu, P. (2006b). Functional nonparametric statistics in action. The art of semi-parametrics 112-129. *Contrib. Statist. Physica- Verlag/Springer, Heidelberg*.
22. Ferraty F., Laksaci, A., and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Processes*, **9**, 47-76.
23. Ferraty F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data : Inference and practical aspects, Austral. *New Zealand J. Statist*, **49**, 267-286.
24. Ferraty F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*, **140**, 335-352.
25. Gao, F., et Li, W.V., (2007). Small ball probabilities for the Slepian Gaussian fields. *Trans. Amer. Math. Soc*, **359** (3) 1339-1350 (electronic).
26. Gasser, T., Hall, P., Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J.Roy. Statist. Soc., Ser B*, **60** :681-691.
27. Györfi, L., Kohler, M., Krzyzak, A., and Walk, H.(2002). A Distribution-Free Theory of Nonparametric Regression. *Springer, New-york*.
28. Konakov, V.D. (1974). On the asymptotic normality of the mode of multidimensional distributions. *Theory of Probab. Appl.*, **19**, 794-799.
29. Laloë, T. (2008). A k-nearest neighbor approach for functional regression. *Statistics & Probability Letters*, **78**, **10**, 1189-1193.
30. Li, W.V., and Shao, Q.M. (2001). Gaussian processes : inequalities, small ball probabilities and applications. In :C.R. Rao and D. Shanbhag (eds.) Stochastic processes, Theory and Methods. *Handbook of Statitics*, **19**, North-Holland, Amsterdam.
31. Lian, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electronic Journal of Statistics*, **5**, 31-40.
32. Lifshits, M.A., et Simon, T. (2005). Small deviations for fractional stable processes. *Ann. Inst. H. Poincaré Probab. Statist*, **41**, 725-752.

33. Lifshits, M.A., Linde, W., et Shi, Z. (2006). Small deviations of Riemann- Liouville processes in Lq-spaces with respect to fractal measures. *Proc. London Math. Soc.*, (3) **92** (1) 224-250.
34. Louani, D., Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametric Statist.*, **11**, 413-442.
35. Lu, Z., and Cheng, P. (1998). Strong consistency of nearest neighbor kernel regression estimation for stationary dependent samples. *Science in China Series A : Mathematics* **41**, **977**, 918-926.
36. Mack, Y.P. (1981). Local properties of kNN regression estimates. *SIAM J. Algebr Discrete Methods*, **2**, 311-323.
37. Ould-Saïd, E. (1993). Estimation non paramétrique du mode conditionnel. Application à la prévision. *C. R. Acad. Sci. Paris. Série I*, **316**, 943-947.
38. Ould-Saïd, E. A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Statist.* **24**, 231-239.
39. Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065-1076.
40. Romano, J.P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.*, **16**, 629-647.
41. Royall, R.M. A class of nonparametric estimates of a smooth regression function, Ph.D. Diss., Stanford University.
42. Samanta, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.*, **7**, 109-117.
43. Samanta, M., and Thavaneswaran, A. (1990). Non-parametric estimation of conditional mode, *Comm. Statist. Theory and Meth.*, **16**, 4515-4524.
44. Shmileva, E. (2006). Small ball probabilities for jump Lévy processes from the Wiener domain of attraction. *Statist. Probab. Lett.*, **76** (17) 1873-1881.
45. Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**.
46. Stone, C.J. (1997). Consistent nonparametric regression, *Ann. Statist.* , **5**, 595-645.
47. Tran, L. T. (1993). Nonparametric Estimation for Time Series by Local Average Estimators. *The Annals of Statistics* **42**, **2**, 1040-1057.

Chapitre 2

Convergence presque complète : cas i.i.d

Ce chapitre fait l'objet d'une publication au journal *Revue Roumaine de Mathématiques Pures et Appliquées*.

The k -nearest neighbour estimation for functional nonparametric conditional mode estimator for functional i.i.d data

Attouch Mohammed Kadi¹ and Bouabça Wahiba²

^{1,2}Univ. Sidi Bel Abbès, Laboratoire de Statistique Processus Stochastiques, Algérie,
attou_kadi@yahoo.fr, wahiba_bouab@yahoo.fr

Abstract

The aim of this article is to study the nonparametric estimation of the conditional mode by the k -Nearest Neighbor method for functional explanatory variables. We give the rate of the almost complete convergence of the k -NN kernel estimator of the conditional mode. A real data application is presented to exhibit the effectiveness of this estimation method with respect to the classical kernel estimation.

Key words : functional data ; the conditional mode, nonparametric regression, k -NN estimator, rate of convergence, random bandwidth.

2.1 Introduction

The conditional mode, by its importance in the nonparametric forecasting field, has motivated a number of researchers in the investigation of mode estimators, and constitutes an alternative method to estimate the conditional regression (see Ould Saïd (1997) for more discussion and examples).

In finite dimension spaces, there exists an extensive bibliography for independent and dependent data cases. In the independent case, strong consistency and asymptotic normality using the kernel method estimation of the conditional mode is given in Samanta and Thavaneswaran (1990). In the dependent case, the strong consistency of conditional mode estimator was obtained by Collomb et al. (1987) and Ould-Saïd (1997) under strong mixing conditions, (see Azzahrioui (2006) for extensive literature).

In infinite dimension spaces, the almost complete convergence of the conditional mode is obtained by Ferraty et al. (2006), Ezzahrioui and Ould Saïd (2008) establish the asymptotic normality of nonparametric estimators of the conditional mode for both independent and dependant functional data. The consistency in L^p -norm of the conditional mode function estimator is given in Dabo-Niang and Laksaci (2007).

Estimating the conditional mode is directly linked to density estimation , and for the last one, the bandwidth selection is extremely important for the performance of an estimate. The bandwidth must not be too large, so as prevent over-smoothing, i.e. substantial bias, and must not be too small either, so as prevent detecting the underlying structure. Particularly, in nonparametric curve estimation, the smoothing parameter is critical for performance.

Based on this fact, this work deals with the nonparametric estimation with k nearest neighbors method k -NN. More precisely we consider a kernel estimator of the conditional mode build from a local window taking into account the exact k nearest neighbors with real response variable Y and functional curves X .

The k -Nearest Neighbor or k -NN estimator is a weighted average of response variables in the neighborhood of x . The existent bibliography of the k -NN method estimation dates back to Royall (1966) and Stone (1977) and has received, continuous developments (Mack (1989) derived the rates of convergence for the bias and variance as well as asymptotic normality in multivariate case, Collomb (1980) studied different types of convergence (probability, almost.sure (a.s) and almost.completely (a.co)) of the estimator of the regression function. Devroye (1978) obtained the strong consistency and the uniform convergence). For the functional data studies, the k -NN kernel estimate was first introduced in the monograph of Ferraty and Vieu (2006), Burba et al. (2009) obtained the rate of almost complete convergence of the regression function using the k -NN method for independent data and Attouch and Benchikh (2012) established the asymptotic normality of robust nonparametric regression function.

The paper is organized as follows : the following section presents the k -NN model estimation of the conditional mode. Then, we give hypotheses and state our main result in Section 3. After we

introduce the almost complete convergence and the almost complete convergence rate, in Section 5 we illustrate the effectiveness of the k -NN method estimation for real data. All proofs are given in the appendix.

2.2 Models and estimators

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n independent pairs, identically distributed as (X, Y) which is a random pair valued in $\mathcal{F} \times \mathbb{R}$, where \mathcal{F} is a semi-metric space, $d(\cdot, \cdot)$ denoting the semi-metric. We do not suppose the existence of a density for the functional random variable (f.r.v) X .

All along the paper when no confusion will be possible we will denote by c, c' or $/ C_x$ some generic constant in \mathbb{R}^+ and in the following, any real function with an integer in brackets an exponent denotes its derivative with the corresponding order. From a sample of independent pair (X_i, Y_i) , each having the same distribution as (X, Y) , our aim is to build nonparametric estimates of several function related with the conditional probability distribution of Y given X . For $x \in \mathcal{F}$ we will denote the conditional distribution function (cdf) of Y given $X = x$ by :

$$\forall y \in \mathbb{R} \quad F_Y^X = \mathbb{P}(Y \leq y | X = x).$$

If this distribution is absolutely continues with respect to the Lebesgues measure on \mathbb{R} , then we will denote by f^x (resp $f^{x(j)}$) the conditional density (resp its j^{th} order derivative) of Y given $X = x$. Then we will give almost complete convergence results (with rates) for nonparametric estimates of $f^{x(j)}$. we will deduce immediately the convergence of conditional density estimate from the general results concerning $f^{x(j)}$.

In the following, x will be a fixed in \mathcal{F} ; N_x will denote a fixed neighborhood of x , \mathcal{S} will be a fixed compact subset of \mathbb{R} , and we will use the notation $B(x, h) = \{x' \in \mathcal{F} / d(x, x') < h\}$.

The conditional distribution in Ferraty et al. (2006) is defined as follows :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))G(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad \forall y \in \mathbb{R}. \quad (2.1)$$

K is a kernel function, G is a conditional distribution function (cdf) and $h = h_K := h_{K,n}$ and $h_G = h_{G,n}$ are sequence of positive real numbers which goes to zero as n goes to infinity.

Under a differentiability assumption of $\widehat{F}^x(y)$, we can obtain the conditional density function by differentiating the conditional distribution function, then we have :

$$\widehat{f}^x(y) = \frac{\partial}{\partial y} \widehat{F}^x(y).$$

We obtain

$$\widehat{f}^x(y) = \frac{h_G^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}. \quad (2.2)$$

$G^{(1)}$ is his first derivative function of G . We assume that $f^x(\cdot)$ has a unique mode of Y , denoted by $\theta(x)$ which is defined by

$$f^x(\theta(x)) = \sup_{y \in \mathcal{S}} f^x(y). \quad (2.3)$$

A kernel estimator of the conditional mode $\theta(x)$ is defined as the random variable $\widehat{\theta}(x)$ which maximizes the kernel estimator $\widehat{f}^x(\cdot)$ of $f^x(\cdot)$

$$\widehat{f}^x(\widehat{\theta}(x)) = \sup_{y \in \mathcal{S}} \widehat{f}^x(y). \quad (2.4)$$

Now we focuses on the estimation of the j^{th} order partial derivative of the conditional density f^x by the k nearest neighbours (k -NN) denoted by $\widehat{f}_{kNN}^{x(j)}(y)$ and defined :

$$\widehat{f}_{kNN}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))G^{(j+1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))}. \quad (2.5)$$

$H_{n,k}(\cdot)$ is defined by :

$$H_{n,k}(x) = \min \left\{ h \in \mathbb{R}^+ / \sum_{i=1}^n \mathbb{I}_{B(x,h)}(X_i) = k \right\}.$$

$H_{n,k}$ is a positive random variable (r.v) which depends (X_1, \dots, X_n) .

A natural $\widehat{\theta}_{kNN}(x)$ is given by :

$$\widehat{f}^x(\widehat{\theta}_{kNN}) = \sup_{y \in \mathcal{S}} \widehat{f}_{kNN}^x(y). \quad (2.6)$$

And it is done directly from equation (2.4), (2.6) and from (2.2) that

$$(\widehat{\theta}_{kNN} - \theta)^j = O \left(\sup_{y \in \mathcal{S}} \left| \widehat{f}_{kNN}^x(y) - f^x(y) \right| \right).$$

Note that this estimate $\widehat{\theta}_{kNN}(x)$ is not necessarily unique, so the remainder of the paper concerns any value $\widehat{\theta}_{kNN}(x)$ satisfying (2.6).

We consider two types of non-parametric models. The first one is called continuity-type, which means that the density function f^x is continuous and the second one, the Hölder-type, which means that the density function f^x is Hölder continuous with constant $b_1, b_2 > 0$.

2.3 Technical tools

The first difficulty comes because the window $H_{n,k}(x)$ is random. So, we do not have in the numerator and in denominator of $f_{kNN}^x(y)$ sums of variables independents. To resolve this problem, the idea is to frame sensibly $H_{n,k}(x)$ by two non-random windows. More generally, these technical tools could be useful as long as one has to deal with random bandwidths.

Lemma 2.3.1 (Burba et al. (2009)). *Let $(A_i, B_i)_{i=1,n}$ be a sequence of random variables, not necessarily identically distributed or independent, with values in $(\omega \times \mathbb{R}, \mathcal{A} \otimes \mathcal{B})$.*

We note $W : \mathbb{R} \times \omega \rightarrow \mathbb{R}^+$ a measurable function such that :

$$(L_0) \quad \forall z \in \omega, \forall t, t' \in \mathbb{R} : \quad t \leq t' \Rightarrow W(t, z) \leq W(t', z).$$

We put for every $n \in \mathbb{N}^$, and T a real random variable (r.r.v)*

$$C_n(T) = \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(T, A_i) h_G^{-1} G_i^{(1)}(h_G^{-1}(y - B_i))}{\sum_{i=1}^n W(T, A_i)} .$$

Let $(D_n)_{n \in \mathbb{N}}$ a sequence of r.r.v and let $(u_n)_{n \in \mathbb{N}}$ a positive decreasing sequence.

1. *if $l = \lim u_n \neq 0$, and if for every increasing sequence $\beta_n \in]0, 1[$, there exists two sequences of r.r.v $D_n^-(\beta_n)$ et $D_n^+(\beta_n)$ which verifies :*

$$(L_1) : \quad \forall n \in \mathbb{N}^* : D_n^- \leq D_n^+, \quad \text{and} \quad \mathbf{1}_{\{D_n^- \leq D_n \leq D_n^+\}} \xrightarrow{\text{a.co.}} 1.$$

$$(L_2) : \quad \frac{\sum_{i=1}^n W(D_n^-, A_i)}{n} \xrightarrow{\text{a.co.}} \beta_n.$$

$$\sum_{i=1}^n W(D_n^+, A_i)$$

$$(L_3) : \quad \exists c \geq 0 \text{ such that : } C_n(D_n^-) \xrightarrow{\text{a.co.}} c \quad \text{and} \quad C_n(D_n^+) \xrightarrow{\text{a.co.}} c .$$

2. *Or if $l = 0$, and if for every increasing sequences $\beta_n \in]0, 1[$ with limits 1, (L_1) , (L_2) and (L_3) are verified.*

So we have :

$$C_n(D_n) \xrightarrow{\text{a.co.}} c .$$

Lemma 2.3.2. (Burba et al. (2009)) *Let $(A_i, B_i)_{i=1,n}$ be a sequence of random variables, not necessarily identically distributed or independent, with values in $(\omega \times \mathbb{R}, \mathcal{A} \otimes \mathcal{B})$.*

We note $W : \mathbb{R} \times \omega \rightarrow \mathbb{R}^+$ a measurable function such that :

$$(L_0) \quad \forall z \in F, \forall t, t' \in \mathbb{R} : \quad t \leq t' \Rightarrow W(t, z) \leq W(t', z).$$

We put for every $n \in \mathbb{N}^*$, and T a r.r.v :

$$C_n(T) = \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(T, A_i) h_G^{-1} G^{(1)}(h_G^{-1}(y - B_i))}{\sum_{i=1}^n W(T, A_i)} .$$

Let $(D_n)_{\mathbb{N}}$ a sequence of r.r.v and let $(v_n)_{n \in \mathbb{N}}$ a positive decreasing sequence.

1. if $l' = \lim v_n \neq 0$, and if for every increasing sequence $\beta_n \in]0, 1[$, there exists two sequences of r.r.v $D_n^-(\beta_n)$ et $D_n^+(\beta_n)$ which verifies :

$$(L'_1) \quad \forall n \in \mathbb{N}^* : D_n^- \leq D_n^+, \quad \text{and} \quad \mathbf{1}_{\{D_n^- \leq D_n \leq D_n^+\}} \xrightarrow{a.co.} 1 ,$$

$$(L'_2) \quad \frac{\sum_{i=1}^n W(D_n^-, A_i)}{n} - \beta_n = O_{a.co.}(v_n) ,$$

$$\sum_{i=1}^n W(D_n^+, A_i)$$

$$(L'_3) \quad \exists c \geq 0 \text{ such that : } C_n(D_n^-) - c = O_{a.co.}(v_n) \text{ and } C_n(D_n^+) - c = O_{a.co.}(v_n) ,$$

2. Or if $l' = 0$, and if for every increasing sequences $\beta_n \in]0, 1[$ with limits 1, (L'_1) , (L'_2) and (L'_3) are verified.

So we have :

$$C_n(D_n) - c = O_{a.co.}(v_n) .$$

Burba et al.(2009) use in their consistency proof of the k -NN kernel estimate for independent data a Chernoff-type exponential inequality to check Condition (L_1) or (L'_1) . In the case of the k -NN conditional mode we can also use the exponential inequality.

Proposition 2.3.1 (Burba et al. (2009)). *Let $X_1 \cdots X_n$ independent random variables valued in $\{0, 1\}$ such as $\mathbb{P}(X_i = 1) = p_i$. We note $X = \sum_{i=1}^n X_i$, and $\mu = \mathbb{E}[X]$, so $\forall \delta > 0$*

1. $\mathbb{P}(X > (1 + \delta)\mu) < \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)$.
2. $\mathbb{P}(X < (1 - \delta)\mu) < e^{-\frac{\delta^2}{2}\mu}$.

Lemma 2.3.3. (Ferraty et al. (2006)). *Under the hypotheses (H1) and (H12) we have :*

$$\frac{1}{n \phi_x(h_K)} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h_K} \right) \xrightarrow{a.co.} 1 .$$

2.4 Hypotheses and results

We need the following hypotheses gathered together for easy references.

(H1) There exists a nonnegative differentiable function ϕ strictly increasing such that :

$$\phi_x(h) = \mathbb{P}(X \in B(x, h)) > 0.$$

(H2) $\forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, |F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2})$
or for some $j \geq 0$.

(H3) $\forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, |f^{x_1(j)}(y_1) - f^{x_2(j)}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2})$.

(H4) $\begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, |G(y_1) - G(y_2)| \leq C|y_1 - y_2| \\ \int |t|^{b_2} G^{(1)}(t) dt < \infty \end{cases}$

(H5) K is a function with support $(0,1)$ such that $0 < C_1 < K(t) < C_2 < \infty$.

(H6) The kernel G is a positive function supported on $[0, 1]$. Its derivative G' exists and is such that there exist two constants C_3 and C_4 with $-\infty < C_3 < G'(t) < C_4 < 0$ for $0 \leq t \leq 1$.

(H7) $\lim_{n \rightarrow \infty} h_G = 0$ with $\lim_{n \rightarrow \infty} n^\alpha h_G = \infty$ for some $\alpha > 0$.

(H8) $\begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, |G^{(j+1)}(y_1) - G^{(j+1)}(y_2)| \leq C|(y_1) - (y_2)|; \\ \exists \nu > 0, \forall j' \leq j + 1, \lim_{y \rightarrow \infty} |y|^{1+\nu} |G^{(j'+1)}(y)| = 0, \\ G^{(j+1)} \text{ is bounded.} \end{cases}$

(H9) $\exists \xi > 0, f^x \nearrow$ on $(\theta - \xi, \theta)$ and $f^x \searrow$ on $(\theta, \theta + \xi)$.

(H10) f^x is j -times continuously differentiable with respect to y on $(\theta - \xi, \theta + \xi)$,

(H11) $\begin{cases} f^{x(l)}(\theta) = 0, \text{ if } 1 \leq l < j, \\ \text{and } |f^{x(j)}(\theta)| > 0. \end{cases}$

(H12) The sequence of positive real numbers $k_n = k$ satisfies : $\frac{k}{n} \rightarrow 0, \frac{\log(n)}{k} \rightarrow 0$ and $\frac{\log(n)}{k h_G^{2j+1}} \rightarrow 0$ as $n \rightarrow \infty$.

Comments on the hypotheses :

1. The assumption (H1) is an assumption commonly used for the explanatory variable X (see Ferraty and Vieu (2006) for more details and examples).
2. Hypotheses (H2)-(H3) characterizes the structural functional space of our model and are needed to evaluate the bias term in our asymptotic properties.
3. (H5) is complementary assumptions concerning the kernel function, focus on non-continuous kernels
4. Hypotheses (H6)-(H8) are the same as those used in Ferraty et al. (2006).
5. The convergence of the estimator can be obtained under the minimal assumption(H9).
6. (H10) and (H11) are classical hypotheses in the functional estimation in finite or infinite dimension spaces.
7. (H12) is technical conditions imposed for the brevity of proofs.

2.5 Almost complete convergence and almost complete convergence rate

In order to link the existing literature with this works, we present the consistency result of the conditional mode estimation given in Ferraty et al. (2006).

Theorem 2.5.1. (Ferraty et al. (2006)) *In the case of a continuity-type, we suppose that hypotheses (H3) and (H8)-(H12) are verified for $j=0$, then if (H1),(H4)-(H5) and (H7)-(H9)-(H11), we have*

$$\lim_{n \rightarrow \infty} \widehat{\theta}(x) = \theta(x) \quad \text{almost completely. (a.co.)}$$

In the case of a Hölder-type model and the same condition of Theorem 2.5.1 we have :

$$\widehat{\theta}(x) - \theta(x) = O\left(h_K^{b_1} + h_G^{b_2}\right)^{\frac{1}{j}} + O\left(\frac{\log(n)}{n h_G \phi_x(h_K)}\right)^{\frac{1}{2j}} \quad \text{a.co.}$$

Now we state the almost complete convergence, and almost complete convergence rate result for the nonparametric k -NN mode introduced in (2.6).

Theorem 2.5.2. *In the case of a continuity-type, suppose that hypotheses (H3) and (H8) are verified for $j=0$, Then if (H1),(H4)-(H5) and (H7)-(H9)-(H12), we have*

$$\lim_{n \rightarrow \infty} \widehat{\theta}_{kNN}(x) = \theta(x) \quad \text{a.co.} \quad (2.7)$$

In the case of a Hölder-type model and the same condition of Theorem 2 we have :

$$\widehat{\theta}_{kNN}(x) - \theta(x) = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^{b_1} + h_G^{b_2}\right)^{\frac{1}{j}} + O\left(\frac{\log(n)}{h_G k}\right)^{\frac{1}{2j}} \quad \text{a.co.} \quad (2.8)$$

Proof

The conditional density $f^x(\cdot)$ is continuous see (H3) and (H9) we get

$$\forall \epsilon > 0, \exists \sigma(\epsilon) > 0, \forall y \in (\theta(x) - \xi, \theta(x) + \xi), |f^x(y) - f^x(\theta(x))| \leq \sigma(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon.$$

By construction $\widehat{\theta}_{kNN}(x) \in (\theta(x) - \xi, \theta(x) + \xi)$ then

$$\forall \epsilon > 0, \exists \sigma(\epsilon) > 0, |f^x(\widehat{\theta}_{kNN}) - f^x(\theta(x))| \leq \sigma(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon.$$

So that we arrive finally at

$$\exists \sigma(\epsilon) > 0, \mathbb{P}(|\widehat{\theta}_{kNN}(x) - \theta(x)| > \epsilon) \leq \mathbb{P}(|f^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| > \delta(\epsilon))$$

In the other case, it comes directly by the definition of $\theta(x)$ and $\widehat{\theta}(x)_{kNN}$ that :

$$\begin{aligned} |f^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| &= |f^x(\widehat{\theta}_{kNN}(x)) - \widehat{f}^x(\widehat{\theta}_{kNN}(x)) + \widehat{f}^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| \\ &\leq |f^x(\widehat{\theta}_{kNN}(x)) - \widehat{f}^x(\widehat{\theta}_{kNN}(x))| + |\widehat{f}^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| \\ &\leq 2 \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)|. \end{aligned} \quad (2.9)$$

The uniform completes convergence of the conditional density estimate over the compact set $|\theta - \xi, \theta + \xi|$ (see Lemma 2.5.3 below) can be used leading directly from both previous inequalities to :

$$\forall \epsilon > 0, \sum_{i=1}^{\infty} \mathbb{P}(|\widehat{\theta}_{kNN}(x) - \theta(x)| > \epsilon) < \infty.$$

Finally, the claimed consistency result (2.7) will be proved as long as the following Lemma could be checked. ■

Lemma 2.5.3. *Under the conditions of Theorem 2.5.2 we have :*

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| = 0 \quad a.co.$$

We have already shown in (2.9) that

$$|f^x(\widehat{\theta}_{kNN}) - f^x(\theta)| = 2 \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)|. \quad (2.10)$$

Let us now write the following Taylor expansion of the function f^x :

$$f^x(\widehat{\theta}_{kNN}) = f^x(\theta) + \frac{1}{j!} f^{x(j)}(\theta^*) (\theta - \widehat{\theta}_{kNN})^j ,$$

for some θ^* between θ and $\widehat{\theta}_{kNN}$ because (2.10), as long as we could be able to check that

$$\exists \tau > 0, \sum_{i=1}^{\infty} \mathbb{P} \left(f^{x(j)}(\theta^*) < \tau \right) < \infty. \quad (2.11)$$

we would have

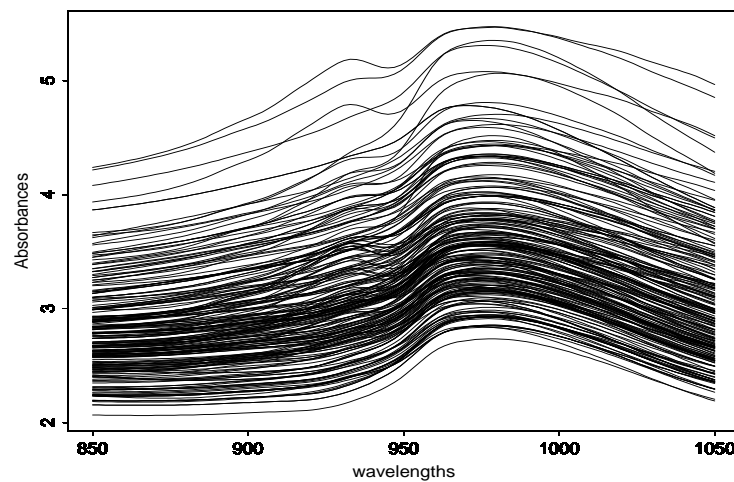
$$(\widehat{\theta}_{kNN} - \theta)^j = O \left(\sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| \right) \quad a.co, \quad (2.12)$$

so it suffice to check (2.11), and this is done directly by using the second part of (H11) together with (2.7). ■

2.6 Real data application

Now we apply the described method to some chemometrical real data. We work on spectrometric data used to classify samples according to a physicochemical property which is not directly accessible (and therefore requires a specific analysis). The data were obtained using a Tecator InfratecFood and Feed Analyzer working in a near-infrared (wavelengths between 850 and 1050 nanometers). The measurement is performed by transmission through a sample finely chopped meat which is then analyzed by a chemical process to determine its fat. The spectra correspond to the absorbance (-log10 of the transmittance measured by the device) 100 for wavelengths between 850 regularly

distributed and 1050 nm. Meat samples were divided into two classes according to their contain more or less than 20% fat (77 was more spectra corresponding to 20% fat and 138 with less than 20% fat). The problem is then to discriminate the spectra to avoid the chemical analysis, which is costly and time consuming. figure shows absorbance versus wavelength (850-1050) for 215 selected pieces of meat. Note that, the main goal of spectrometric analysis is to allow for the discovery of the proportion of some specific chemical content (see Ferraty and Vieu (2006) for further details related to spectrometric data). At this stage, one would like to use the spectrometric curve X to predict Y the proportion of fat content in the piece of meat. The data are available on the web site "<http://lib.stat.cmu.edu/datasets/teacator>".



The 215 spectrometric curves, $\{X_i(t), t \in [850, 1050], i = 1, \dots, 215.\}$

However, as described in the real data application application (see Ferraty et al. (2006)), the prediction problem can be studied by using the conditional mode approach. Starting from this idea, our objective is to give a comparative study to estimate the conditional mode of the both methods : the kernel method estimation defined in (2.4) and k -NN method estimation defined in (2.6).

For the both methods, we use a quadratic kernel function K , defined by :

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{[0,1]},$$

and the distribution function $G(\cdot)$ is defined by

$$G(x) = \int_{-\infty}^x \frac{15}{4}t^2(1 - t^2)\mathbb{I}_{[-1,1]}(t)dt.$$

Note that, the shape of these spectrometric curves is very smooth, for this we used the semi-metric defined by the L^2 distance between the second derivative of the curves (see Ferraty and Vieu (2002) for more motivations of this choice). We proceed by the following algorithm :

- *Step 1.* We split our data into two subsets :
 - $(X_j, Y_j)_{j=1, \dots, 172}$ training sample,
 - $(X_i, Y_i)_{i=173, \dots, 215}$ test sample.
- *Step 2.* We compute the mode estimator $\hat{\theta}_{X_j}$, for all j by using the training sample.
- *Step 3.* For each X_i in the test sample, we set : $i_* = \arg \min_{j=1, \dots, 172} d(X_i, X_j)$.
- *Step 4.* For all $i = 173, \dots, 215$ we take

$$\hat{Y}_i = \hat{\theta}(X_{i_*}),$$

where X_{i_*} is the nearest curve to X_i in the training sample.

And

$$\tilde{Y}_i = \hat{\theta}_{k_{NN}}(X_{k_{opt}}),$$

whereas the optimal number of neighbors k_{opt} is defined by

$$k_{opt} = \arg \min_k CV(k),$$

where

$$CV(k) = \sum_{i=171}^{215} (Y_i - \hat{\theta}_{k_{NN}}^{-i}(x))^2,$$

with

$$\hat{\theta}_{k_{NN}}^{-i}(x) = \arg \max_y (\hat{f}_{k_{NN}}^x)^{-i}(y),$$

and

$$(\hat{f}_{k_{NN}}^x)^{-i}(y) = \frac{h_G^{-1} \sum_{j=1, j \neq i}^{172} K(H_{n,k}(x)^{-1}(x)d(x, X_j))G^{(1)}(h_G^{-1}(y - Y_j))}{\sum_{j=1, j \neq i}^{172} K(H_{n,k}^{-1}(x)d(x, X_j))}.$$

The error used to evaluate this comparison is the mean square error (MSE) expressed by

$$\frac{1}{45} \sum_{i=173}^{215} |Y_i - \hat{T}(X_i)|^2,$$

where \hat{T} designate the estimator used : kernel or k -NN method to estimate the conditional mode. The MSE of k -NN mode estimation is 1.716115 and the MSE of the classical mode estimation used in Ferraty et al.(2006) is 2.97524, this result exhibe the effectiveness of the k -NN method.

2.7 Appendix

Proof of Lemma 2.3.2

For $i=1, \dots, n$, we consider the quantities $G_i^{(1)}(y) = G^{(1)}(h_G^{-1}(y - B_i))$.

Under the definition of the r.v D_n , we put : $D_n^- \leq D_n \leq D_n^+$.

It's clear that :

$$W(D_n^-(\beta_n), A_i) \leq W(D_n(\beta_n), A_i) \leq W(D_n^+(\beta_n), A_i),$$

$$\sum_{i=1}^n W(D_n^-(\beta_n), A_i) \leq \sum_{i=1}^n W(D_n(\beta_n), A_i) \leq \sum_{i=1}^n W(D_n^+(\beta_n), A_i).$$

So :

$$\frac{1}{\sum_{i=1}^n W(D_n^+(\beta), A_i)} \leq \frac{1}{\sum_{i=1}^n W(D_n(\beta_n), A_i)} \leq \frac{1}{\sum_{i=1}^n W(D_n^-(\beta_n), A_i)}.$$

Under the hypotheses (H7) – (H8) and (H9), we have :

$$\sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n^-(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\underbrace{\sum_{i=1}^n W(D_n^+(\beta_n), A_i)}_{C_n^+(\beta_n)}} \leq \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\underbrace{\sum_{i=1}^n W(D_n(\beta_n), A_i)}_{C_n(D_n)}} \leq \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n^+(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\underbrace{\sum_{i=1}^n W(D_n^-(\beta_n)(\beta_\epsilon), A_i)}_{C_n^-(\beta_n)}}.$$

Let us put the following r.v :

$$C_n^-(\beta_n) = \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n^-(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\sum_{i=1}^n W(D_n^+(\beta_n), A_i)}, \quad \text{and} \quad C_n^+(\beta_n) = \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n^+(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\sum_{i=1}^n W(D_n^-(\beta_n), A_i)}.$$

We consider for every $\epsilon > 0$:

$$T_n(\epsilon) = \{ C_n(D_n) / c - \epsilon \leq C_n(D_n) \leq c + \epsilon \},$$

and for every $\beta \in]0, 1[$, we note :

$$S_n^-(\epsilon, \beta_n) = \{ C_n^-(\beta_n) / c - \epsilon \leq C_n^-(\beta_n) \leq c + \epsilon \},$$

$$S_n^+(\epsilon, \beta_n) = \{ C_n^+(\beta_n) / c - \epsilon \leq C_n^+(\beta_n) \leq c + \epsilon \},$$

and

$$S_n(\beta_n) = \{ C_n(D_n) / C_n^-(\beta_n) \leq C_n(D_n) \leq C_n^+(\beta_n) \}.$$

It is obvious that :

$$S_n(\beta_n) \cap S_n^-(\epsilon, \beta_n) \cap S_n^+(\epsilon, \beta_n) \subset T_n(\epsilon). \quad (2.13)$$

Let

$$\epsilon_0 = \frac{3c}{2} \quad \text{and} \quad \beta_\epsilon = 1 - \left(\frac{\epsilon}{3c} \right).$$

By denoting :

$$W_n^-(\epsilon) = \{ C_n^-(\beta_\epsilon) / \beta_\epsilon c - \frac{\epsilon}{3} \leq C_n^-(\beta_\epsilon) \leq \beta_\epsilon c + \frac{\epsilon}{3} \},$$

$$W_n^+(\epsilon) = \{ C_n^+(\beta_\epsilon) / \frac{c}{\beta_\epsilon} - \frac{\epsilon}{3} \leq C_n^+(\beta_\epsilon) \leq \frac{c}{\beta_\epsilon} + \frac{\epsilon}{3} \},$$

$$W_n(\epsilon) = \{D_n / D_n^-(\beta_\epsilon) \leq D_n \leq D_n^+(\beta_\epsilon)\}.$$

We see that for $\epsilon \in]0, \epsilon_0[$:

$$\begin{cases} \beta_\epsilon c - \frac{\epsilon}{3} \geq c - \epsilon, \\ \beta_\epsilon c + \frac{\epsilon}{3} \leq c + \epsilon, \end{cases} \quad \text{and} \quad \begin{cases} \frac{c}{\beta_\epsilon} - \frac{\epsilon}{3} \geq c - \epsilon, \\ \frac{c}{\beta_\epsilon} + \frac{\epsilon}{3} \leq c + \epsilon. \end{cases}$$

So, for $\epsilon \in]0, \epsilon_0[$:

$$W_n^-(\epsilon) \subset S_n^-(\epsilon, \beta_\epsilon) \quad \text{and} \quad W_n^+(\epsilon) \subset S_n^+(\epsilon, \beta_\epsilon). \quad (2.14)$$

(L_0) implies that :

$$\forall z \in F, \forall t, t' \in \mathbb{R} : \quad t \leq t' \Rightarrow W(t, z) \leq W(t', z).$$

And in particular :

$$\begin{aligned} D_n \in W_n(\epsilon) &\Rightarrow D_n^-(\beta_\epsilon) \leq D_n \leq D_n^+(\beta_\epsilon), \\ D_n \in W_n(\epsilon) &\Rightarrow C_n^-(\beta_\epsilon) \leq C_n(D_n) \leq C_n^+(\beta_\epsilon), \\ &\Rightarrow C_n(D_n) \in S_n(\beta_\epsilon). \end{aligned}$$

So :

$$W_n(\epsilon) \subset S_n(\beta_\epsilon). \quad (2.15)$$

The result (2.13), (2.14) and (2.15) implies :

$$\forall \epsilon \in]0, \epsilon_0[, \quad W_n^-(\epsilon) \cap W_n^+(\epsilon) \cap W_n(\epsilon) \subset T_n(\epsilon).$$

In other words :

$$\forall \epsilon \in]0, \epsilon_0[, \quad \mathbb{1}_{cT_n(\epsilon)} \leq \mathbb{1}_{cW_n^-(\epsilon)} + \mathbb{1}_{cW_n^+(\epsilon)} + \mathbb{1}_{cW_n(\epsilon)}. \quad (2.16)$$

In the other hand, we can express the r.r.v C_n^- and C_n^+ in the following way :

$$\begin{aligned} C_n^-(\beta_n) &= \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n^-(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\sum_{i=1}^n W(D_n^+(\beta_n), A_i)}, \\ &= C_n(D_n^-) \times \frac{\sum_{i=1}^n W(D_n^-(\beta_n), A_i)}{\sum_{i=1}^n W(D_n^+(\beta_n), A_i)}, \end{aligned}$$

and

$$\begin{aligned}
 C_n^+(\beta_n) &= \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(D_n^+(\beta_n), A_i) h_G^{-1} G_i^{(1)}(y)}{\sum_{i=1}^n W(D_n^-(\beta_n), A_i)}, \\
 &= C_n(D_n^+) \times \frac{\sum_{i=1}^n W(D_n^+(\beta_n), A_i)}{\sum_{i=1}^n W(D_n^-(\beta_n), A_i)}.
 \end{aligned}$$

So, under (L_2) and (L_3) :

$$C_n^-(\beta_n) \xrightarrow{a.co.} \beta_n c, \quad \text{and} \quad C_n^+(\beta_n) \xrightarrow{a.co.} c/\beta_n. \quad (2.17)$$

Because :

$${}^c W_n^-(\epsilon) = \{C_n^-(\beta_\epsilon) / |C_n^-(\beta_\epsilon) - \beta_\epsilon c| > \frac{\epsilon}{3}\},$$

the first parte of (2.17) implies :

$$\begin{aligned}
 C_n^-(\beta_\epsilon) \xrightarrow{a.co.} \beta_\epsilon c &\Rightarrow \forall \epsilon' > 0 : \sum \mathbb{P}(|C_n^-(\beta_\epsilon) - \beta_\epsilon c| > \epsilon') < \infty, \\
 &\Rightarrow \forall \epsilon \in]0, \epsilon_0[: \sum \mathbb{P}(\mathbb{1}_{{}^c W_n^-(\epsilon)} > \epsilon) < \infty.
 \end{aligned}$$

So :

$$\forall \epsilon \in]0, \epsilon_0[, \mathbb{1}_{{}^c W_n^-(\epsilon)} \xrightarrow{a.co.} 0. \quad (2.18)$$

In the same way :

$$\forall \epsilon \in]0, \epsilon_0[, \mathbb{1}_{{}^c W_n^+(\epsilon)} \xrightarrow{a.co.} 0. \quad (2.19)$$

Under (L_1) : $\mathbb{1}_{\{D_n^- \leq D_n \leq D_n^+\}} \xrightarrow{a.co.} 1$,

it implies that : $\forall \epsilon \in]0, \epsilon_0[: \mathbb{1}_{W_n(\epsilon)} \xrightarrow{a.co.} 1$.

In the other word :

$$\forall \epsilon \in]0, \epsilon_0[, \mathbb{1}_{{}^c W_n(\epsilon)} \xrightarrow{a.co.} 0. \quad (2.20)$$

We consider the result (2.16), (2.18), (2.19), and (2.20) implies :

$$\forall \epsilon \in]0, \epsilon_0[, \mathbb{1}_{{}^c T_n(\epsilon)} \xrightarrow{a.co.} 0.$$

That we can repeat as follows :

$$\begin{aligned}
 \mathbb{1}_{{}^c T_n(\epsilon)} \xrightarrow{a.co.} 0 &\Rightarrow \forall \epsilon \in]0, \epsilon_0[: \sum \mathbb{P}(|\mathbb{1}_{{}^c T_n(\epsilon)}| > \epsilon) < \infty, \\
 &\Rightarrow \forall \epsilon \in]0, \epsilon_0[: \sum \mathbb{P}(|C_n(D_n) - c| > \epsilon) < \infty.
 \end{aligned}$$

And we obtain finally :

$$C_n(D_n) \xrightarrow{a.co.} c.$$

■

Proof of Theorem 2.5.2 (Almost complete convergence)

 Let $G_i^{(1)}(y) = G^{(1)}(h_G^{-1}(y - Y_i))$.

For the demonstration of this Theorem we have to use Lemma 2.3.2 with :

- $\forall i = 1, n : (A_i, B_i) = (X_i, Y_i)$.
- $D_n = H_{n,k}$, and $C_n(D_n) = \sup_{y \in \mathcal{S}} \hat{f}_{kNN}^x(y)$.
- $\forall (t, z) \in \mathbb{R}^+ \times E : W(t, z) = K\left(\frac{d(x, z)}{t}\right)$.

 And for one $\beta_n \in]0, 1[$, we choose D_n^- et D_n^+ Such that :

$$\phi_x(D_n^-) = \sqrt{\beta_n} \frac{k}{n} \quad \text{and} \quad \phi_x(D_n^+) = \frac{1}{\sqrt{\beta_n}} \frac{k}{n}.$$

 For the demonstration, we just have to show that the kernel K verify (L_0) , and that the sequences $(D_n^-)_{\mathbb{N}}$, $(D_n^+)_{\mathbb{N}}$ et $(D_n)_{\mathbb{N}}$ verify (L_1) , (L_2) and (L_3) .

Verification of (L_0) :

$$\begin{aligned} \forall t, t' \in \mathbb{R}^+ : t \leq t' &\Leftrightarrow \exists v \in [1, +\infty[/ t' = vt, \\ &\Leftrightarrow \exists u \in]0, 1] / \frac{1}{t'} = u \frac{1}{t}, \\ &\Leftrightarrow \exists u \in]0, 1] / K\left(\frac{\|x - z\|}{t'}\right) = K\left(u \frac{\|x - z\|}{t}\right) \quad \forall z \in E. \end{aligned}$$

 We have that K is a bounded kernel, verifying : $\forall z \in E, \forall u \in [0, 1] : K(uz) \geq K(z)$ so we have :

$$K\left(\frac{\|x - z\|}{t}\right) \leq K\left(u \frac{\|x - z\|}{t}\right),$$

where :

$$\forall t, t' \in \mathbb{R}^+, \forall z \in E : t \leq t' \Rightarrow K\left(\frac{\|x - z\|}{t}\right) \leq K\left(\frac{\|x - z\|}{t'}\right).$$

 So (L_0) is verified.

Verification of (L_1) :

 The proof of (L_1) is the same one as in Burba et al. (2009).

Verification of (L_2) :

 According to $(H12)$:

$$\begin{cases} \lim_{n \rightarrow +\infty} \phi_x(D_n^-) = \lim_{n \rightarrow +\infty} \sqrt{\beta_n} \frac{k}{n} = 0, \\ \lim_{n \rightarrow +\infty} \phi_x(D_n^+) = \lim_{n \rightarrow +\infty} \frac{1}{\sqrt{\beta_n}} \frac{k}{n} = 0. \end{cases}$$

 For $(H1)$, we deduce that :

$$\begin{cases} \lim_{n \rightarrow +\infty} D_n^- = 0, \\ \lim_{n \rightarrow +\infty} D_n^+ = 0. \end{cases}$$

And by (H12) we have :

$$\left\{ \begin{array}{l} \lim_{n \rightarrow +\infty} \left(\frac{\log(n)}{n \phi_x(D_n^-)} \right) = \lim_{n \rightarrow +\infty} \left(\frac{\log(n)}{n \sqrt{\beta_n} \left(\frac{k}{n} \right)} \right) = 0, \\ \lim_{n \rightarrow +\infty} \left(\frac{\log(n)}{n \phi_x(D_n^+)} \right) = \lim_{n \rightarrow +\infty} \left(\frac{\sqrt{\beta_n} \log(n)}{n \left(\frac{k}{n} \right)} \right) = 0. \end{array} \right.$$

Because (D_n^-) and (D_n^+) verify the conditions of Lemma 2.3.3, we have :

$$\left\{ \begin{array}{l} Q(D_n^-) = \frac{1}{n \phi_x(D_n^-)} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) \xrightarrow{a.co.} 1, \\ Q(D_n^+) = \frac{1}{n \phi_x(D_n^+)} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) \xrightarrow{a.co.} 1. \end{array} \right.$$

So :

$$\frac{Q(D_n^-)}{Q(D_n^+)} = \frac{\left(\frac{1}{n \phi_x(D_n^-)} \right) \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)}{\left(\frac{1}{n \phi_x(D_n^+)} \right) \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} \xrightarrow{a.co.} 1.$$

We can deduce that :

$$\frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} \xrightarrow{a.co.} \beta_n.$$

So, (L_2) is checked.

Verification of (L_3) :

Because D_n^- et D_n^+ verify the condition of Theorem 2.5.1, we obtain :

$$\left\{ \begin{array}{l} \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) h_G^{-1} G_i^{(1)}(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)} \xrightarrow{a.co.} f^x(y), \\ \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) h_G^{-1} G_i^{(1)}(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} \xrightarrow{a.co.} f^x(y). \end{array} \right.$$

So, (L_3) is checked.

What finishes the proof of the first part of Theorem 2.5.2. ■

Proof of Theorem 2.5.2 (Almost complete convergence rate)

In the same way as the demonstration of the first part of Theorem 2.5.2 (almost complete convergence), with the following notations :

- $\forall i = 1, n : (A_i, B_i) = (X_i, Y_i)$
- $D_n = H_n$, et $C_n(D_n) = \hat{f}_{kNN}^x(y)$
- $\forall (t, z) \in \mathbb{R}^+ \times E : W(t, z) = K\left(\frac{d(x, z)}{t}\right)$.

And for one $\beta_n \in]0, 1[$, we choose D_n^- and D_n^+ such that :

$$\phi_x(D_n^-) = \sqrt{\beta_n} \frac{k}{n}, \quad \text{and} \quad \phi_x(D_n^+) = \frac{1}{\sqrt{\beta_n}} \frac{k}{n}.$$

For the proof, it is enough to show that the kernel K verify (L_0) , and the sequences $(D_n^-)_{\mathbb{N}}$, $(D_n^+)_{\mathbb{N}}$ and $(D_n)_{\mathbb{N}}$ Verify (L_1) , (L'_2) and (L'_3) .

Verification of (L_0) and (L_1) :

The verification of (L_0) and (L_1) is made in the same way such that for Lemma 2.3.1.

Verification of (L'_2) :

We have already shown in Ferraty et al (2006) that :

$$\hat{f}_N^x(y) = \frac{1}{n h_G \mathbb{E}[K_1]} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)) G_i^{(1)}(y), \quad \text{and} \quad \hat{F}_D^x = \frac{1}{n \mathbb{E}[K_1]} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)).$$

So :

$$\hat{f}^x(y) - f^x(y) = \frac{1}{\hat{F}_D^x} \left\{ \left(\hat{f}_N^x(y) - \mathbb{E} \hat{f}_N^x(y) \right) - \left(f^x(y) - \mathbb{E} \hat{f}_N^x(y) \right) \right\} + \frac{\hat{f}_N^x(y)}{\hat{F}_D^x} \left\{ \mathbb{E} \hat{F}_D^x - \hat{F}_D^x \right\}.$$

To show that :

$$\sup_{y \in \mathcal{S}} \left| \hat{f}^x(y) - f^x(y) \right| = O(h_K^{b_1}) + O(h_G^{b_2}) + O\left(\sqrt{\frac{\log n}{n h_G \phi_x(h_K)}}\right) \quad a.co.$$

It is enough to show that :

$$(A) \quad \frac{1}{\hat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \mathbb{E} \hat{f}_N^{x(j)}(y) - f^{x(j)}(y) \right| = O(h_K^{b_1}) + O(h_G^{b_2}), \quad a.co.$$

$$(B) \quad \frac{1}{\hat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \hat{f}_N^{x(j)}(y) - \mathbb{E} \hat{f}_N^{x(j)}(y) \right| = O\left(\sqrt{\frac{\log(n)}{n h_G^{2j+1} \phi_x(h_K)}}\right), \quad a.co.$$

$$(C) \quad \hat{F}_D^x - \mathbb{E} \hat{F}_D^x = O\left(\sqrt{\frac{\log(n)}{n \phi_x(h_K)}}\right), \quad a.co.$$

$$(D) \quad \exists \delta > 0, \quad \text{such that} \quad \sum \mathbb{P} \left[\left| \hat{F}_D^x \right| < \delta \right] < +\infty.$$

We put : $h^- = D_n^- = \phi_x^{-1} \left(\sqrt{\beta_n} \frac{k}{n} \right)$, and $h^+ = D_n^+ = \phi_x^{-1} \left(\frac{1}{\sqrt{\beta_n}} \frac{k}{n} \right)$.

Because : $\lim_{n \rightarrow +\infty} h^- = \lim_{n \rightarrow +\infty} h^+ = 0$,

and $\lim_{n \rightarrow +\infty} \frac{\log(n)}{n h_G \phi_x(D_n^-)} = \lim_{n \rightarrow +\infty} \frac{\log(n)}{n h_G \phi_x(D_n^+)} = \lim_{n \rightarrow +\infty} \frac{\log(n)}{h_G k} \lim_{n \rightarrow +\infty} \frac{\log(n)}{k} = 0$, because β_n is bounded by 1).

We apply Lemma 2.3.3, and we obtain the both result :

1. $\frac{1}{n} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h^-} \right) - \phi_x(h^-) = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{n \phi_x(h^-)}} \right)$.

What implies :

$$\frac{1}{n} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) - \sqrt{\beta_n} \frac{k}{n} = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{k}} \right) .$$

2. $\frac{1}{n} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h^+} \right) - \phi_x(h^+) = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{n \phi_x(h^+)}} \right)$.

So :

$$\frac{1}{n} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) - \frac{1}{\sqrt{\beta_n}} \frac{k}{n} = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{k}} \right) .$$

We have :

$$\frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} - \beta_n = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{k}} \right) .$$

And according to (H12)

$$\frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} - \beta_n = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{h_G k}} \right) .$$

(L'_2) is also verified.

Verification of (L'_3) :

Because h^+ and h^- verify the conditions, we can apply Theorem 2.5.1, what gives :

$$\left\{ \begin{array}{l} \left| \frac{\sup_{y \in \mathcal{S}} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) h_G G_i^{(1)}(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)} - \sup_{y \in \mathcal{S}} f^x(y) \right| = O(D_n^-)^{b_1} + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{n h_G \phi_x(D_n^-)}} \right) \\ \left| \frac{\sup_{y \in \mathcal{S}} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) h_G G_i^{(1)}(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} - \sup_{y \in \mathcal{S}} f^x(y) \right| = O(D_n^+)^{b_1} + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{n h_G \phi_x(D_n^+)}} \right) \end{array} \right.$$

In the other word :

$$\left\{ \begin{array}{l} \left| C_n(D_n^-) - \sup_{y \in \mathcal{S}} f^x(y) \right| = O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} \right) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{h_G k}} \right) \\ \left| C_n(D_n^+) - \sup_{y \in \mathcal{S}} f^x(y) \right| = O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} \right) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{h_G k}} \right) \end{array} \right.$$

So : $\exists c = \sup_{y \in \mathcal{S}} f^x(y) > 0$, such that :

$$\left\{ \begin{array}{l} |C_n(D_n^-) - c| = O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{h_G k}} \right) \\ |C_n(D_n^+) - c| = O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{h_G k}} \right) \end{array} \right.$$

So (L'_3) is verified .

We have finally : $\sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| = O \left(\phi_X^{-1} \left(\frac{k}{n} \right)^\alpha \right) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{\log(n)}{h_G k}} \right)$.

■

2.8 Conclusion and perspectives

We proposed in this work a convergence results of the nonparametric k -NN method estimation of the conditional mode for functional data. The almost complete pointwise convergence of this estimator with rates is given.

The effectiveness of the k -NN method resort on practical examples, as long as the smoothing parameter k takes values in a discrete set this provides a better implementation.

This work can be generalized to the dependent data (see Azzahrioui (2006)). In nonparametric statistics, uniform convergence is considered as a preliminary step to obtain sharper results, in this context, it would be very interesting to extend the results of Ferraty et al. (2010).

Obtaining explicit expressions of the dominant terms of centered moments can be envisaged when we obtain the asymptotic normal result (see, Delsol (2008)). This idea can be investigate in the future.

Acknowledgments. The authors thanks a Referee, the Editors-in-chief for helpful comments.

Références

1. Attouch, M., and Benchikh, T. (2010). Asymptotic distribution of robust k -nearest neighbour estimator for functional nonparametric models. *Mathematic Vesnic*, **64**, no. 4, 275-285.
2. Bosq, D. (2000). Linear processes in function spaces. Theory and Application, Lectures Notes in Statistics. *Springer Verlag, New-York*, Vol. **149**.
3. Burba, F., Ferraty, F., and Vieu, P. (2008). Convergence of the k nearest neighbor kernel estimator in nonparametric functional regression. *C. R. Math. Acad. Sci. Paris*, **346**, no. 5-6, 339-342.
4. Burba, F., Ferraty, F., and Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.*, **21**, no. 4, 453-469.
5. Collomb, G. (1980). Estimation de la régression par la méthode des k points les plus proches avec noyau : quelques propriétés de convergence ponctuelle, in Nonparametric Asymptotic Statistics. *Lecture notes in mathematics, Springer-Verlag*, **821**, 159-175.
6. Collomb, G., Härdle, W., and Hassani, S. (1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf*, **15**, 227-236.
7. Dabo-Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie : Application aux diffusions. *C. R. Acad. Sci., Paris*, **334**, No.3, 213-216, 2002.
8. Dabo-Niang, S. (2004). Kernel density estimator in an infinite dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett*, **17** :381- 386, 2004.
9. Dabo-Niang, S., Ferraty, F., and Vieu, P. (2006). Mode estimation for functional random variable and its application for curves classification. *Far East J. Theoretical Statist*, **18**, 93-119.
10. Dabo-Niang, D., and A. Laksaci. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C. R., Math., Acad. Sci. Paris*, **334**, 49-52.
11. Delsol, L. (2008). Advances on asymptotic normality in nonparametric functional time series analysis. *Statistics*, **43**, No. 1, 13-33.
12. Devroye, L.P. (1978). The uniform convergence of nearest neighbour regression function estimators and their application in optimization. *IEEE Trans. Inform. Theory*, **24**, 142-151.
13. Eddy, W. F. (1982). The asymptotic distribution of kernel estimators of the mode. *Z. W. Giebeta*, **59**, 279-290.
14. Ezzahrioui, M., and Ould-Saïd, E. (2006). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. Preprint, LMPA No 277, Univ. du Littoral Côte d'Opale.
15. Ezzahrioui, M., and Ould-Saïd, E. (2008). Asymptotic normality of nonparametric estimator of the conditional mode for functional data. *Journal of Nonparametric Statistics*, **20**, 3-18.

16. Ferraty, F., and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist. and Data Anal*, **4** :545-564.
17. Ferraty, F., and Vieu, P. (2006). Nonparametric Functional Data Analysis : Theory and practice. *Springer series in statistics, Springer. New York*.
18. Ferraty, F., Laksaci, A., and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Processes*, **9**, 47-76.
19. Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data : Inference and practical aspects, Austral. *New Zealand J. Statist*, **49**, 267-286.
20. Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables, *Journal of Statistical Planning and Inference*, **140**, 335-352.
21. Gasser, T., Hall, P., Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc., Ser B*, **60** :681-691.
22. Mack, Y.P. (1981). Local properties of k NN regression estimates. *SIAM J. Algebr Discrete Methods*, **2**, 311-323.
23. Ould-Saïd, E. (1993). Estimation nonparamétrique du mode conditionnel. Application à la prévision. *C. R. Acad. Sci. Paris. Série I*, **316**, 943-947.
24. Ould-Saïd, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Statist*, **24**, 231-239.
25. Royall, R.M. (1996). A class of nonparametric estimates of a smooth regression function, Ph.D. Diss., Stanford University.
26. Samanta, M., and Thavaneswaran, A. (1990). Non-parametric estimation of conditional mode. *Comm. Statist. Theory and Meth.*, **16**, 4515-4524.
27. Stone, C.J. (1997). Consistent nonparametric regression. *Ann. Statist*, **5**, 595-645.

Chapitre 3

Normalité asymptotique : Cas i.i.d

Ce travail traite la normalité asymptotique lorsque les données sont fonctionnelles i.i.d. Ce travail fait l'objet d'un article soumis.

Some asymptotic normality result of k -Nearest Neighbour estimator of the conditional mode function for independent functional data

Attouch Mohammed Kadi¹ and Bouabça Wahiba²

^{1,2}Univ. Sidi Bel Abbès, Laboratoire de Statistique Processus Stochastiques, Algérie,
attou_kadi@yahoo.fr, wahiba_bouab@yahoo.fr

Abstract This paper, deal with the nonparametric estimation of the conditional mode using the k -Nearest Neighbors (k-NN) method for a scalar response variable given a random variable taking values in a semi-metric space. We establish the asymptotic normality for independent functional data, and propose confidence bands for the conditional mode function. Some simulations have been driven to show how our methodology can be implemented.

Key words : functional data ; k-NN estimator ; the conditional mode estimation ; small balls probability ; asymptotic normality.

3.1 Introduction

The conditional mode, by its importance in the nonparametric forecasting field, has motivated a number of researchers in the investigation of mode estimators, and constitutes an alternative method to estimate the conditional regression.

The motivation for this mode regression model is its interest in some nonparametric estimation problems where the mode regression provides better estimations than the classical mean regression (see for instance Collomb et al. (1987) Quintela and Vieu (1997), Ould-Saïd (1997), Berlinet et al. (1998), or Louani and Ould-Saïd (1999), for the multivariate case).

The statistical problems involved in the modelization of functional random variables has received increasing interests in recent literature (see for example Dabo-Niang (2002), Ferraty and Vieu (2004), Dabo-Niang and Rhomari (2003), Masry (2005) for nonparametric context).

In this functional area, the first results concerning the conditional mode estimation were obtained by Ferraty et al. (2006a). They established the almost complete convergence of the kernel estimator in the i.i.d. case. This last result has been extended to dependent case by Ferraty et al (2005), Ezzahrioui and Ould-saïd (2006a, 2006b) have studied the asymptotic normality of the kernel estimator of the conditional mode for both i.i.d. and strong mixing cases. The monograph of Ferraty and Vieu (2006b) presents an important collection of statistical tools for nonparametric prediction of functional variables. Recently, Dabo-Niang and Laksaci(2007) stated the convergence in L^p norm of the condition al mode function in the independent case.

The k -Nearest Neighbour estimator is a weighted average of response variables in the neighborhood

of x , this estimation method take into account the local structure of the data. The k -NN kernel estimate has a significant advantage over the classical kernel estimate. The main advantage of the k -NN method is in the nature of the smoothing parameter. Indeed, in the classical kernel method, the smoothing parameter is the bandwidth h_n , which is a real positive number and in k -NN method, the smoothing parameter k_n takes its values in discrete set.

The k -NN kernel estimate is a widely studied estimate if the explanatory variable is an element of a finite-dimensional space, see Györfi et al (2002). In the functional case with real-valued response, two different approaches for the k -NN regression estimation exist. The first one, published by Laloë (2008), examines a k -NN kernel estimate when the functional variable is an element of a separable Hilbert space. For that case Laloë establishes a weak consistency result. However, his ansatz is not completely functional. Laloë's strategy is to reduce the dimension of the input variable by using a projection onto a finite-dimensional subspace and then applying multivariate techniques on the projected data. The second result, from Burba et al (2009), is based on a pure functional approach instead. Burba et al. examine the problem on a semi-metric functional space. They proved almost complete convergence and rates for independent data. Furthermore, Burba et al. extended a lemma that we will also use in our proofs. This lemma originates from Collomb (1980). Additionally, the k -NN kernel estimate is examined for classification in infinite dimension by Cérou and Guyader (2006) and there exists a convergence result for the k -NN regression estimate when the response is an element of a Hilbert space (see Lian (2011)).

The first result of the asymptotic normality on the kernel estimator comes from Masry (2005), he considers the case of α -mixing data , but he did not give the explicit expression of the dominant asymptotic terms of bias and variance. After Ferraty et al.(2007) gives the explicit expression of the asymptotic law (that is the dominant terms of bias and variance) In the case of set of independent data . The results exposed in the papers of Delsol (2007a, 2008b) make the link between these articles, they generalize the results of Ferraty et al. (2007) in the case of data α -mixing, Attouch and Benchikh (2012) established the asymptotic normality of robust nonparametric regression function. For the k -NN conditional mode estimator Attouch and Bouabça (2013) obtained the almost complete convergence with rates in independent and identically distributed (i.i.d.) functional data case.

3.2 The model

Let $(X_i, Y_i), i = 1, \dots, n$ be n copies of random vector identically distributed as (X, Y) where X is valued in infinite dimensional semimetric vector space (\mathcal{F}, d) and Y 's are valued in \mathbb{R} , In most practical applications, \mathcal{S} is a normed space which can be of infinite dimension (e.g. Hilbert or Banach space) with norm $\|\cdot\|$ so that $d(x, x') = \|x - x'\|$.

We will denote the conditional distribution function of Y by :

$$\forall y \in \mathbb{R} \quad F^x(y) = \mathbb{P}(Y \leq y | X = x).$$

Then we will denote by f^x (resp $f^{x(j)}$) the conditional density (resp its j^{th} order derivative) of Y . Furthermore we will give almost complete convergence results (with rates) for nonparametric estimates of $f^{x(j)}$. Since $f^x = f^{x(0)}$, we will deduce immediately the convergence of conditional density estimate from the general results concerning $f^{x(j)}$.

The conditional distribution in Ferraty et al.(2006b) is defined as follows :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))G(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad \forall y \in \mathbb{R}. \quad (3.1)$$

K is a kernel function, G is a distribution function (df) and $h = h_K := h_{K,n}$ and $h_G = h_{G,n}$ are sequence of positive real numbers which goes to zero as n goes to infinity.

Under a differentiability assumption of $G(\cdot)$, we can obtain the conditional density estimation function as :

$$\widehat{f}^x(y) = \frac{h_G^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}, \quad (3.2)$$

and its j^{th} partial derivative with respect to y as :

$$\widehat{f}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(j)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}. \quad (3.3)$$

$G^{(j)}$ is the j^{th} order derivative function of G . We assume that $f^x(\cdot)$ has a unique mode, denoted by $\theta(x)$ which is defined by

$$f^x(\theta(x)) = \sup_{y \in \mathcal{S}} f^x(y). \quad (3.4)$$

A kernel estimator of the conditional mode $\theta(x)$ is defined as the random variable $\widehat{\theta}(x)$ which maximizes the kernel estimator $\widehat{f}^x(\cdot)$ of $f^x(\cdot)$

$$\widehat{f}^x(\widehat{\theta}(x)) = \sup_{y \in \mathcal{S}} \widehat{f}^x(y). \quad (3.5)$$

Let consider quantities

$$\widehat{f}^x(y) = \frac{\widehat{f}_N^x(y, h)}{\widehat{F}_D^x(h)}, \quad (3.6)$$

where

$$\widehat{f}_N^x(y, h) = \frac{1}{n h_G \phi(h)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(1)}(h_G^{-1}(y - Y_i)), \quad (3.7)$$

$$\widehat{F}_D^x(h) = \frac{1}{n \phi(h)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)), \quad (3.8)$$

where $\phi(\cdot)$ is a function supposed to be strictly positive and it will be described later.

An analogous estimator to (3.7) was already given in Ferraty et al. (2006a) in the general setting, remark that (3.7) and (3.8) are unbiased estimators of $a_1 f^x(y) F_D^x(h) = a_1 f_N^x(y, h)$ and $a_1 F_D^x$ respectively where a_1 and F_D^x will be described later.

Now we focus on the estimation of the j^{th} order derivative of the conditional density f on x by the k nearest neighbors (k -NN), where $\widehat{f}_{kNN}^{x(j)}(y)$ is defined by :

$$\widehat{f}_{kNN}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i)) G^{(j+1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))}. \quad (3.9)$$

$H_{n,k}(\cdot)$ is defined by :

$$H_{n,k}(x) = \min \left\{ h \in \mathbb{R}^+ / \sum_{i=1}^n \mathbb{I}_{B(x,h)}(X_i) = k \right\}.$$

$H_{n,k}$ is a positive random variable (r.v) which depends (X_1, \dots, X_n) .

From now on, when we refer to the bandwidth of the k -NN conditional mode estimation, we mean the number of neighbours k we are considering.

A natural $\widehat{\theta}_{kNN}(x)$ is given by :

$$\widehat{f}^x(\widehat{\theta}_{kNN}) = \sup_{y \in \mathbb{R}} \widehat{f}_{kNN}^x(y). \quad (3.10)$$

so that

$$\widehat{f}_{kNN}^x(y) = \frac{\widehat{f}_N^x(y, H_{n,k})}{\widehat{F}_D^x(H_{n,k})}. \quad (3.11)$$

Note that this estimate $\widehat{\theta}_{kNN}(x)$ is not necessarily unique, so the remainder of the paper concerns any value $\widehat{\theta}_{kNN}(x)$ satisfying (3.10).

We consider two types of non-parametric models. The first one is called continuity-type, which means that the density function f is continuous and the second one, the Hölder-type, which means that the density function f is Hölder continuous with constant $b_1, b_2 > 0$.

Our main goal is to establish the asymptotic normality of the estimator in (3.10) when suitably normalized. As a consequence we get the asymptotic normality of a predictor and propose confidence bands for the conditional mode function with the k -NN method. The paper is organized as follows : in Section 2 we give the assumptions and main results. After we introduce the asymptotic normality in Section 3, some applications are given in Section 4. Finally, the proof of the main result is relegated to Section 5 with some auxiliary results.

3.3 Hypotheses and results

We give some additional notations to formulate our assumptions, let $B(x, h)$ be the ball of center x and radius h and $M_i = \|x - X_i\|$ a random variable such that $\mathbb{P}(X_i \in B(x, h)) = \mathbb{P}(M_i \leq h) = W_x(h)$ and for all fixed $x \in \mathcal{S}$. Furthermore let \mathcal{S} be a compact set of \mathbb{R} such that $\theta(x) \in \overset{\circ}{\mathcal{S}}$, where $\overset{\circ}{\mathcal{S}}$ denotes the interior of \mathcal{S} .

(H1) There exist three functions $F_D^x(\cdot)$, $\phi(\cdot)$ (supposed increasing and strictly positive for every $h > 0$) and $\zeta_0(\cdot)$ such that

1. $W_x(h) = F_D^x(h)\phi(h) + O(\phi(h))$.
2. $\forall u \in [0, 1], \lim_{h \rightarrow 0} \frac{\phi(uh)}{\phi(h)} = \lim_{h \rightarrow 0} \zeta_h(u) = \zeta_0(u)$.

(H2) The kernel K is nonnegative, with compact support $[0, 1]$ of class C^1 on $[0, 1]$, $K(0) > 0$ and $K(1) > 0$ and its derivative K' exists on $[0, 1]$ and $K'(t) < 0$.

(H3) The conditional density and its two first derivatives satisfy the Hölder condition

$$\forall (y_1, y_2) \in \mathbb{R}^2, \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, |f^{x_1(j)}(y_1) - f^{x_2(j)}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}).$$

for $j = 0, 1, 2$ where \mathcal{N}_x is a neighborhood of x , and C_x is constant that depends on x .

(H4) $G^{(1)}(\cdot)$ is twice differentiable such that

1. $\int_{\mathbb{R}} |t^2|G^{(1)}(t)dt < +\infty$, and $\int_{\mathbb{R}} tG^{(1)}(t)dt = 0$,
2. $\int_{\mathbb{R}} (G^{(2)}(t))^2 dt < +\infty$, $\int_{\mathbb{R}} |t^{b_2}|(G^{(2)}(t))^2 dt < +\infty$,
3. $\forall (y_1, y_2) \in \mathbb{R}^2, |G^{(j)}(y_1) - G^{(j)}(y_2)| \leq C|y_1 - y_2|$ for $j = 1, 3$ and $G^{(j)}$ are bounded for $j = 1, 2, 3$.

(H5) The bandwidth h_G satisfies, for $0 \leq j \leq 2$

1. $\frac{\log^2(n)}{nh_G^{j+2}\phi_x(h)} \rightarrow 0$, as $n \rightarrow \infty$,
2. $n^\beta h_G^{j+2} \rightarrow +\infty$ as $n \rightarrow \infty$ for some $\beta > 0$.

(H6) The sequence of positive real numbers $k_n = k$ satisfies :

$$\frac{k}{n} \rightarrow 0 \text{ and } \frac{\log(n)}{k} \rightarrow 0, \frac{\log(n)}{k h_G^{2j+1}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Remarks on the assumptions

1. The assumption (H1) is classical for the explanatory variable X . this assumption replaces the condition of a strict positivity of the density of explanatory variable usually assumed in finite-dimensional case. The decomposition of the concentration measure as product of two independent functions has been adopted in Masry (2005) who has reformulated the Gasser et al. (1998) and used after by many authors (see Attouch et al. (2009) for comment and some explicit examples).

The function $\zeta_h(\cdot)$ is increasing for all fixed h . Its pointwise limiting $\zeta_0(\cdot)$ will also play a

determinant role. It will intervene in all asymptotic behaviour, It permits to give the variance term explicitly. We quote the following exemples which can be found in Ferraty et al. (2007) :

- i) $\zeta_0(u) = u^\gamma$,
- ii) $\zeta_0(u) = \delta_1(u)$ where $\delta_1(u)$ is Dirac's function,
- iii) $\zeta_0(u) = \mathbb{I}_{]0,1[}(u)$.

2. Assumptions (H2) , (H4) and (H5) are classical in the functional estimation in finite or infinite dimension spaces. Recall that assumption (H4.1) implies that (H1) has finite moment of order.
3. Hypotheses (H3) characterize the structural functional space of our model and is needed to evaluate the bias term in our asymptotic properties.
4. Condition (H6) is technical condition imposed for the brevity of proofs.

By the definition of the conditional mode function we have

$$f^{x(1)}(\theta(x)) = 0.$$

Similarly, it follows that

$$\widehat{f}^{x(1)}(\widehat{\theta}_{kNN}(x)) = 0.$$

Furthermore, we assume that $f^{x(2)}(\theta(x)) \neq 0$, and $\widehat{f}^{x(2)}(\widehat{\theta}_{kNN}(x)) \neq 0$.

By a Taylor's expansion of $\widehat{f}^{x(1)}(\cdot)$ in the neighborhood of $\theta(x)$, we get

$$\widehat{f}^{x(1)}(\widehat{\theta}_{kNN}(x), u) = \widehat{f}^{x(1)}(\theta(x), u) + (\widehat{\theta}_{kNN}(x) - \theta(x))\widehat{f}^{x(2)}(\theta^*(x), u),$$

with $\widehat{\theta}^*(x) \in (\widehat{\theta}_{kNN}(x), \theta(x))$. Because of the definition $\widehat{\theta}_{kNN}(x)$, we have

$$\widehat{\theta}_{kNN}(x) - \theta(x) = -\frac{\widehat{f}^{x(1)}(\theta(x), H_{n,k})}{\widehat{f}^{x(2)}(\widehat{\theta}^*(x), H_{n,k})}; \quad (3.12)$$

using equation (3.11) and (3.9) we can write

$$\widehat{\theta}_{kNN}(x) - \theta(x) = -\frac{\widehat{f}_N^{x(1)}(\theta(x), H_{n,k})}{\widehat{f}_N^{x(2)}(\widehat{\theta}^*(x), H_{n,k})}; \quad (3.13)$$

if the denominator does not vanish. In what follows, we use the notations $f_N^{x(j)}(y) = F_D^x f^{x(j)}(y)$ for $j = 0, 1, 2$.

Our main result is given in the following theorem.

Theorem 3.3.1. *Assume that (H1) and (H6) hold, then for any $x \in \Upsilon$, we have*

$$\left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} (\widehat{\theta}_{kNN}(x) - \theta(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \quad (3.14)$$

$\xrightarrow{\mathcal{D}}$ means the convergence in distribution, $\Upsilon = \{x : x \in \mathcal{S}, f^x(\theta(x)) \neq 0\}$

$$\sigma^2(x, \theta(x)) = a_2 f^x(\theta(x)) \int_{\mathbb{R}} (G^{(2)}(t))^2 dt \quad (\text{with } a_j = K^j(1) - \int_0^1 (K^j(s))' \zeta_0(s) ds, \text{ for } j = 1, 2). \quad (3.15)$$

The detailed proof of Theorem 3.3.1 is postponed to appendix.

3.4 Some perspectives for predictive confidence band

3.4.1 Prediction

For $n \in \mathbb{N}^*$, let $X_i(t), i = 1, \dots, n$ be functional random variables with $t \in \mathbb{R}$. For each curve $X_i(t)$, we have also at hand the value of some real response variables Y_i which corresponds to some modality of our problem.

The aim important question which is posed in force in practice is the following one : given a new curve $X_{n+1}(t) = x_{new}$, can we predict the corresponding response y_{new} ?.

This is a typical prediction problem for infinite-dimensional explanatory random variables.

The predictor estimator is defined by computing the quantity

$$\hat{y}_{new} = \hat{\theta}_{kNN}(x_{new}) = \arg \max_y \hat{f}_{kNN}^{x_{new}}(y).$$

The theoretical predictor of x_{new} is clearly defined by equation (3.10) that is

$$y_{new} = \theta(x_{new}).$$

Applying Theorem 3.3.1, we have corollary 3.4.1.

Corollary 3.4.1. *Under the hypotheses of Theorem 3.3.1 we have*

$$\left(\frac{h_G^3 k(a_1 f^{x_{new}(2)}(\theta(x_{new})))^2}{\sigma^2(x_{new}, \theta(x_{new}))} \right)^{1/2} \left(\hat{\theta}_{kNN}(x_{new}) - \theta(x_{new}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \quad (3.16)$$

3.4.2 Predictive confidence bands

Our main application of the above Theorem is the to build confidence interval for the true value of θ given curve $X = x$. Let us define consistent estimators of the normalization constants a_l for $l = 1, 2$, that is

$$\hat{a}_l = \frac{1}{h_G k} \sum_{i=1}^n K^l \left(\frac{d(x, X_i)}{\phi^{-1}(k/n)} \right).$$

This estimation is justified by the fact that, under (H1) and (H8), we have, (see Ferraty and Vieu (2006b, p. 44)

$$\frac{1}{h_G k} \mathbb{E}[K_1^l] \rightarrow a_l, \quad l = 1, 2.$$

A plug-in estimate for the asymptotic standard deviation $\sigma(x, \theta(x)) = (Var(x, \theta(x)))^{1/2}$, can be obtained using the estimators $\hat{\theta}_{kNN}(x)$, $\hat{f}_{kNN}^{(\cdot)}$ and \hat{a}_2 of $\theta(x)$, $f^{(\cdot)}$ and a_2 , respectively, that is

$$\hat{\sigma}(x, \hat{\theta}_{kNN}(x)) = \left(\hat{a}_2 \hat{f}^x(\hat{\theta}_{kNN}(x)) \int_{\mathbb{R}} (G^{(2)}(t))^2 dt \right)^{\frac{1}{2}}. \quad (3.17)$$

Clearly, $\hat{\sigma}(x, \hat{\theta}_{kNN}(x))$ is a consistent estimator of $\sigma(x, \theta(x))$. Then using the above estimator and the estimates \hat{a}_1 and $\hat{f}^{(2)}$ of a_1 and $f^{(2)}$ respectively, yields the following asymptotic $(1 - \zeta)$

confidence bands for $\theta(x)$.

$$\widehat{\theta}_{kNN}(x) \pm \mu_{1-\zeta/2} \times \left(\frac{\widehat{\sigma}^2(x, \widehat{\theta}_{kNN}(x))}{h_G^3 k \left(\widehat{a}_1 \widehat{f}^{x(2)}(\widehat{\theta}_{kNN}(x)) \right)^2} \right)^{1/2}, \quad (3.18)$$

where $\mu_{1-\zeta/2}$ denotes the $(1 - \zeta/2)$ quantile of the standard normal distribution.

Finally, the approximate $(1 - \zeta/2)$ confidence band, for any $x \in \Upsilon$, is

$$[\varrho_-(x), \varrho_+(x)] \text{ where } \varrho_{\pm}(x) = \widehat{\theta}_{kNN}(x) \pm \mu_{1-\zeta/2} \times \left(\frac{h_G^2 \sum_{i=1}^n K_i^2 \widehat{f}^x(\widehat{\theta}_{kNN}(x))}{\left(\sum_{i=1}^n K_i \widehat{f}^{x(2)} \widehat{\theta}_{kNN}(x) \right)^2} \right)^{1/2}.$$

Here we point out that, due to simplification, the function $\phi(\cdot)$ does not intervene in the practical point of view.

3.5 Simulation study

In this section we will show the effectiveness of k -NN method compared to the kernel estimation using simulated data. For this we considered a sample of a diffusion process on interval $[0, 1]$, $Z_1(t) = 2 - \cos(\pi t W)$ and $Z_2(t) = \cos(\pi t W)$, where W is the standard normal distribution and take $X(t) = AZ_1(t) + (1-A)Z_2(t)$, where A is random variable Bernoulli distributed. We carried out the simulation with a 200 sample of the curve X which is represented by the following graph :

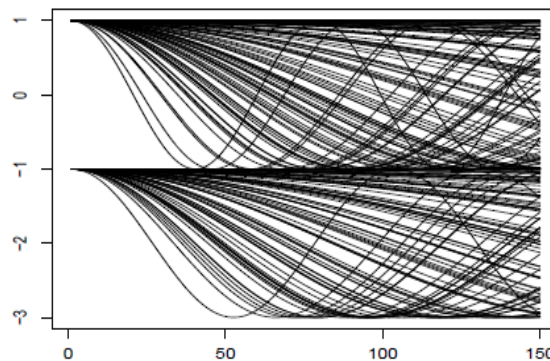


FIGURE 3.1 – The 200 curves X .

For the scalar response variable, we define $Y = Ar_1(X) + (1 - A)r_2(X)$ where r_1 (resp. r_2) is the nonlinear regression model $r_1(X) = 0.25 \times \left(\int_0^1 X'(t) dt \right)^2 + \epsilon$, with ϵ is $U([0, 0.5])$ (resp. $r_2(X)$ is

the null function). We choose a quadratic kernel K defined by :

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{[0,1]}, \quad \text{and } K(1) > 0. \quad (3.19)$$

and the distribution function $G(\cdot)$ is defined by

$$G(x) = \int_{-\infty}^x \frac{15}{4}t^2(1 - t^2)\mathbb{I}_{[-1,1]}(t)dt. \quad (3.20)$$

In practice, the semi-metric choice is based on the regularity of the curves X . For this we use the semi-metric defined by the L_2 -distance between the q^{th} derivatives of the curves. In order to evaluate the MSE (Mean Square Error) we proceed by the following algorithm :

- *Step 1.* We split our data into two subsets ; the first sample, of size $n=120$ corresponds to the learning sample which will be used, as a sample, to compute our conditional mode function estimators for the 80 remaining curves (considered as the test sample).
 - $(X_j, Y_j)_{j \in J}$ learning sample,
 - $(X_i, Y_i)_{i \in I}$ test sample.
- *Step 2.* We use the learning sample for computing the hazard function estimator $\hat{\theta}_{X_j}$, for all $j \in J$.
- *Step 3.* For each X_i in the test sample, we set : $i_* = \arg \min_{j \in J} d(X_i, X_j)$.
- *Step 4.* We put : $\forall i \in I$ and we take

$$\hat{Y}_i = \hat{\theta}(X_{i_*}),$$

where X_{i_*} is the nearest curve to X_i in the training sample, and

$$\tilde{Y}_i = \hat{\theta}_{k_{NN}}(X_{k_{opt}}),$$

whereas the optimal number of neighbours k_{opt} is defined by

$$k_{opt} = \arg \min_k CV(k),$$

where

$$CV(k) = \sum_{i=121}^{200} (Y_i - \hat{\theta}_{k_{NN}}^{-i}(x))^2,$$

with

$$\hat{\theta}_{k_{NN}}^{-i}(x) = \arg \max_y (\hat{f}_{k_{NN}}^x)^{-i}(y),$$

and

$$(\hat{f}_{k_{NN}}^x)^{-i}(y) = \frac{h_G^{-1} \sum_{l=1, l \neq i}^{120} K(H_{n,k}^{-1}(x)d(x, X_l))G^{(1)}(h_G^{-1}(y - Y_l))}{\sum_{l=1, j \neq i}^{120} K(H_{n,k}^{-1}(x)d(x, X_l))}.$$

The error used to evaluate this comparison is the mean square error (MSE) expressed by

$$\frac{1}{\text{card}(I)} \sum_{i \in I} |Y_i - \hat{T}(X_i)|^2,$$

where \widehat{T} designate the estimator used : kernel or k -NN method to estimate the conditional mode. The MSE of k -NN mode estimation is 0.991851 and the MSE of the classical mode estimation used in Ferraty et al. (2006) is 15.55373.

- *Step 4.* We implement now our methodology and see how good the asymptotic normality in the determination of the confidence bands.

We define the confidence bands for all $i \in I$ by :

$$\left[\widehat{\theta}(X_{i^*}) \pm \mu_{0.975} \times \left(\frac{\widehat{\sigma}^2(X_{i^*}, \widehat{\theta}(X_{i^*}))}{h_G^3 k \left(\widehat{a}_1 \widehat{f}^{X_{i^*}(2)}(\widehat{\theta}(X_{i^*})) \right)^2} \right) \right]$$

where $\mu_{0.975}$ is the 5% quantile of a standard normal distribution.

We present our results by plotting the extremities of the predicted values versus the true values in the Figure 3.2.

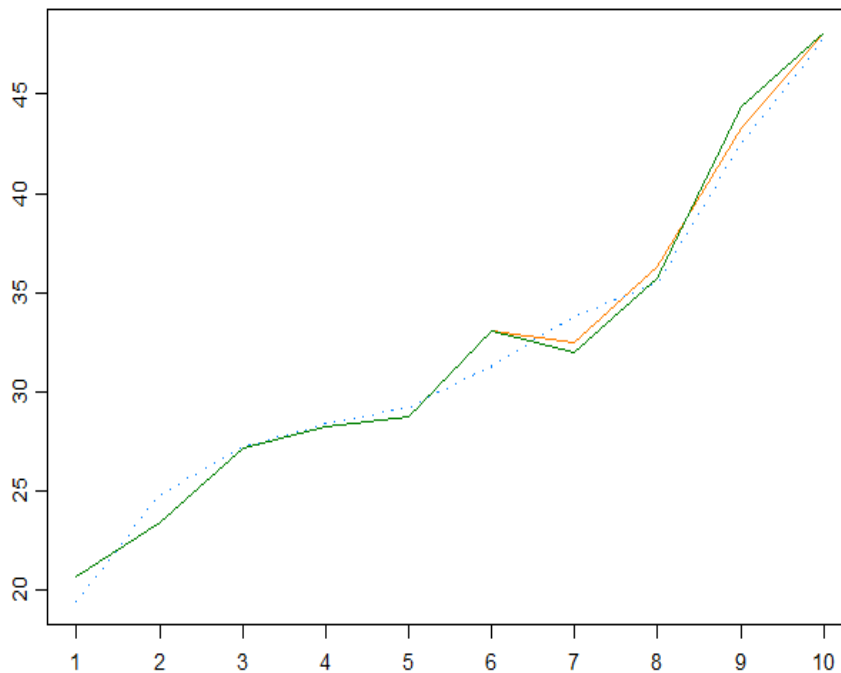


FIGURE 3.2 – the 95% conditional predictive bands.

3.6 Appendix

Using equation (3.13), with $D_n = H_{n,k}$, $c_n(D_n) = \widehat{f}_N^x(y, h)$, $c = f^x(y)$ and let $\lambda \in]0, 1[$, we choose D_n^- and D_n^+ as

$$W_x(D_n^-) = \sqrt{\lambda} \frac{k}{n}, \quad W_x(D_n^+) = \frac{1}{\sqrt{\lambda}} \frac{k}{n},$$

we have

$$\widehat{\theta}_{kNN}(x) - \theta(x) = \frac{-c_n^{(1)}(\theta(x), H_{n,k})}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})}. \quad (3.21)$$

Then, we obtain this decomposition :

$$\begin{aligned} \left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} (\widehat{\theta}_{kNN}(x) - \theta(x)) &= \left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \left(\frac{-c_n^{(1)}(\theta(x), H_{n,k})}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})} \right) \\ &= \left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \left[\frac{-c_n^{(1)}(\theta(x), H_{n,k}) + c_n^{(1)}(\theta(x), D_n^+) - c_n^{(1)}(\theta(x), D_n^+)}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})} \right] \\ &= - \left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \frac{c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})} \\ &\quad + \left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \frac{-c_n^{(1)}(\theta(x), H_{n,k}) + c_n^{(1)}(\theta(x), D_n^+)}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})} \\ &\quad - \left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \frac{\mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})}. \end{aligned} \quad (3.22)$$

Then, to state asymptotic normality, we show that the numerator of the first term of the right hand side of (3.22) suitably normalized is asymptotically normally distributed, the numerator of the last term goes to zero in probability as n goes to infinity and the denominator converges in probability to $(a_1(x) f^{x(2)}(\theta(x)))$.

For this, we remind that, under the same assumptions as Theorem 2.1, Ezzahrioui et al. (2007) proved that

$$\left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \left(c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \quad (3.23)$$

Lemma 2.2 of Ezzahrioui et al. (2007) gives that

$$\left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \left(\mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] \right) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (3.24)$$

In the other hand, by hypothesis (H2), and the fact that $\mathbb{I}_{\{D_n^- \leq H_{n,k} \leq D_n^+\}} \xrightarrow{a.co.} 1$ when $\frac{k}{n} \rightarrow 0$ (see Burba et al.(2009)), we have

$$c_n^{(1)}(\theta(x), D_n^+) \leq c_n^{(1)}(\theta(x), H_{n,k}) \leq c_n^{(1)}(\theta(x), D_n^-).$$

Using that

$$\begin{aligned}
 \left| c_n^{(1)}(\theta(x), H_{k,n}) - c_n^{(1)}(\theta(x), D_n^+) \right| &\leq \left| c_n^{(1)}(\theta(x), D_n^-) - c_n^{(1)}(\theta(x), D_n^+) \right| \\
 &\leq \left| c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] \right| \\
 &\quad + \left| \mathbb{E}[\theta(x), c_n^{(1)}(\theta(x), D_n^+)] - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] \right| \\
 &\quad + \left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-) - c_n^{(1)}(\theta(x), D_n^-)] \right|.
 \end{aligned} \tag{3.25}$$

The equation

$$\left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \left| c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] \right|, \tag{3.26}$$

is the same as (3.23) and Lemma 2.3 of Ezzahrioui et al (2007) gives that

$$\left(\frac{h_G^3 k (a_1 f^{x(2)}(\theta(x)))^2}{\sigma^2(x, \theta(x))} \right)^{1/2} \left| c_n^{(1)}(\theta(x), D_n^-) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] \right| \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{3.27}$$

For the third term, we have :

$$\left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] \right| \leq \left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] - c \right| + \left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] - c \right|. \tag{3.28}$$

The almost complete convergence to 0 of these two terms is verified in (3.26) and (3.27).

On the other hand, to establish the convergence in probability of denominator in (3.22), note that

$$\left| c_n^{(2)}(\hat{\theta}^*(x), H_{n,k}) - a_1(x) f^{x(2)}(\theta(x)) \right| \leq \left| c_n^{(2)}(\hat{\theta}^*(x), D_n^+) - c_n^{(2)}(\theta(x), D_n^+) \right| + \left| c_n^{(2)}(\theta(x), D_n^+) - a_1(x) f^{x(2)}(\theta(x)) \right|. \tag{3.29}$$

By Lemma 2.3 and Lemma 2.4 in Ezzahrioui et al. (2007) and the continuity of $c_n^{(2)}(\cdot, \cdot)$ we deduce that the first term of (3.29) converges in probability to 0. Furthermore, this result in conjunction with the Lemma 2.1 improves the result of Ferraty et al. (2006a).

However, the limit of second term is consequence of the following inequality and the Lemma 2.3 of Ezzahrioui et al. (2007)

$$\left| c_n^{(2)}(\theta(x), H_{n,k}) - a_1(x) f^{x(2)}(\theta(x)) \right| \leq \left| c_n^{(2)}(\theta(x), D_n^+) - a_1(x) f^{x(2)}(\theta(x)) \right|.$$

■

Acknowledgements The authors would like to thank the anonymous reviewers, whose remarks helped us improve substantially the quality of the paper and clarify some points. Many thanks to one of the two referees whose permit us to improve the presentation of the article.

Références

1. Attouch M., Bouabça, W. (2013). The k -nearest neighbors estimation of the conditional mode for functional data. *Rev. Roumaine Math. Pures Appl*, **58**, No. 4, pp. 393-415.
2. Attouch M., Benchikh, T. (2012). Asymptotic distribution of robust k -nearest neighbour estimator for functional nonparametric models. *Mathematic Vesnic*, **64**, No. 4, pp. 275-285.
3. Attouch, M., Laksaci, A., Ould-Saïd, E. (2009). Asymptotic distribution of robust estimator for functional nonparametric models. *Comm. Statist. Theory and Methods*, **38 (8)**, 1317-1335.
4. Attouch M., Bouabça, W. (2013). The k -nearest neighbors estimation of the conditional mode for functional data. *Rev. Roumaine Math. Pures Appl*, **58**, No. 4, pp. 393-415.
5. Burba, F., Ferraty, F., Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression. *J. Nonparametric. Stat*, **21(4)**, 453-469.
6. Berlinet, A., Gannoun, A. Matzner-Lober, E. (1998). Normalité asymptotique d'estimateurs convergents du mode conditionnel. *Rev. Canad. Statist*, **26**, 365-380.
7. Collomb, G., Härdle, W. and Hassani, S. (1987). A note on prediction via conditional mode estimation, *J. Statist. Plann. and Inf*, **15**, 227-236.
8. Collomb, G. (1980). Estimation de la regression par la methode des k points les plus proches avec noyau : quelques propriétés de convergence ponctuelle. In *Statistique non Paramétrique Asymptotique*, J.-P. Raoult, Ed., vol. **821** of *Lecture Notes in Mathematics*, Springer Berlin /Heidelberg, pp. 159-175.
9. Cérou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM : Probability and Statistics*. **10**, 340-355.
10. Dabo-Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie : application aux diffusions. "*C. R., Math., Acad. Sci. Paris*", **334**, pp. 213-216.
11. Dabo-Niang, S. and Rhomari, N. (2003). Estimation non paramétrique de la regression avec variable explicative dans un espace métrique. *C. R., Math., Acad. Sci. Paris*, **336**, No.1, 75-80.
12. Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C. R., Math., Acad. Sci. Paris*, **344**, 49-52.
13. Delsol, L. (2007a). CLT and L^q errors in nonparametric functional regression. *C. R. Math. Acad. Sci*, **345 (7)** 411-414.
14. Delsol, L. (2008b). Advances on asymptotic normality in nonparametric functional Time Series Analysis Statistics.
15. Ezzahrioui, M., Ould-Saïd, E. (2006a). Asymptotic normality of the kernel estimators of the conditional quantile in the normed space. *LMPA* n 258.
16. Ezzahrioui, M., Ould-Saïd, E. (2006b). Asymptotic results of the kernel estimator of the conditional quantile in the normed space under α -mixing hypothesis.

17. Ezzahrioui, M., Ould-Saïd, E. (2007). Prédiction dans les modèles conditionnels en dimension finie, PhD Thesis.
18. Ezzahrioui, M. and Ould-Saïd, E. (2010). Some Asymptotic result of nonparametric estimator of the conditional mode for functional time-series data. *Journal of Statistica Neerlandica*, **64**, 171-201.
19. Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination. *J. Nonparametr. Stat.*, **v16**. 11-127.
20. Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. *C. R., Math., Acad. Sci. Paris*, **340**, No.5, 389-392.
21. Ferraty, F., Laksaci, A. and Vieu, P. (2006a). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Proc*, **9**, No.2, 47-76.
22. Ferraty, F. and Vieu, Ph. (2006b). Nonparametric functional data analysis. *Springer-Verlag, New- York*.
23. Ferraty, F., Mas, A., Vieu, P. (2007). Nonparametric regression on functional data : Inference and practical aspects, *Australian and New Zealand Journal of Statistics*, **49**, 267-286.
24. Gasser, T., Hall, P., Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc., Ser B*, **60** :681-691.
25. Györfi, L., Kohler, M., Krzyzak, A., and Walk, H. (2002). A Distribution-Free Theory of Nonparametric Regression. *Springer, New-york*.
26. Lian, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electronic Journal Of Statics*, **5**, 31-40.
27. Laloë, T. (2008). A k-nearest neighbor approach for functional regression. *Statistics & Probability Letters*, **78**, 10, 1189-1193.
28. Louani, D. and Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametric Stat*, **11** (4), 413-442.
29. Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Process. Appl*, **115**, No.1, 155-177.
30. Ould-Saïd, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Stat*, **24**, No.2, 231-239.
31. Quintela del Rio, A. and Vieu, Ph. (1997). A nonparametric conditional mode estimate, *Nonparametric Statistics*, **8**, 253-266.

Chapitre 4

Convergence presque complète : cas α -mélange

Ce travail traite la convergence presque complète du mode conditionnel lorsque les données sont fonctionnelles dépendantes. Ce travail fait l'objet d'un article soumis.

The k -nearest neighbour estimation for functional nonparametric conditional mode estimator for functional time-series data**Attouch Mohammed Kadi¹ and Bouabça Wahiba² and Chiker El Mezoaur Zouaoui³**^{1,2}Univ. Sidi Bel Abbès, Laboratoire de Statistique Processus Stochastiques, Algérie,

attou_kadi@yahoo.fr, wahiba_bouab@yahoo.fr

³Université de Béchar, Algérie, chikrtime@yahoo.fr

Abstract. In this paper, the k -Nearest Neighbours method estimation (shortly k -NN) of the conditional mode is investigated for scalar response variable Y given a functional random variable X . We give asymptotic results of the k -NN conditional mode estimation : the almost complete convergence and its rate under strongly mixing dependence condition. To exhibit the effectiveness of this estimation method compared to the classical kernel method estimation a real data application is given.

4.1 Introduction

There exist many instances where the conditional density of the regression function becomes zero anywhere and therefore can not be used in some problems implying prediction. Starting from this point of view, Ould Saïd (1997) in finite dimensional spaces and Ferraty et al. (2006) in infinite dimensional spaces show that the conditional mode provides an alternative approach to the classical regression function.

Nonparametric estimation for dependent observations has a long history in statistics. Rosenblatt (1970) first studied density estimation for dependent data. Since then several authors have considered nonparametric estimation under various assumptions (notable early articles include Robinson (1983) and Hart (1984)). For example, Giraitis et al. (1996), Mielniczuk (1997) and Estevas and Vieu (2003) consider density estimation for linear processes which have long memory, whereas Cheng and Robinson (1991) consider density estimation for random variables which are nonlinear functions of a linear process.

In finite dimension spaces, there exists an extensive bibliography for dependent data cases. The strong consistency of conditional mode estimator was obtained by Collomb et al. (1987) and Ould-Saïd (1997) under strong mixing conditions. (See Azzahrioui and Ould-Saïd (2006) for extensive literature).

In infinite dimension spaces, the almost complete convergence (a.co.) of the conditional mode is obtained by Ferraty et al. (2006), Ezzahrioui and Ould Saïd (2008) establish the asymptotic normality of nonparametric estimators of the conditional mode for both independent and dependant functional data. The consistency in L^p -norm of the conditional mode function estimator is given in Dabo-Niang and Laksaci (2007).

Let $(X_i, Y_i), i = 1, \dots, n$ be n copies of random vector identically distributed as (X, Y) where X is valued in infinite dimensional semimetric vector space (\mathcal{F}, d) and Y 's are valued in \mathbb{R} . We do not suppose the existence of a density for the functional random variable (f.r.v) X .

For fixed value $x \in \mathcal{F}$; \mathcal{N}_x denote a fixed neighborhood of x , \mathcal{S} is a fixed compact subset of \mathbb{R} and $\theta(x) \in \overset{\circ}{\mathcal{S}}$ where $\overset{\circ}{\mathcal{S}}$ denotes the interior of \mathcal{S} .

All along the paper, when no confusion will be possible, we will denote by C or/and C_x some generic constant in \mathbb{R}^{*+} , and in the following, any real function with an integer in brackets as exponent denotes its derivative with the corresponding order.

Our aim is to build nonparametric estimates of several functions related to the conditional probability distribution of Y given X .

We denote the conditional probability distribution function of Y_i given $X_i = x$ by :

$$\forall y \in \mathbb{R}, \quad F^x(y) = \mathbb{P}(Y \leq y | X = x)$$

which is supposed to be independent of i .

Then we will denote by f^x (resp $f^{x(j)}$) the conditional density (resp its j^{th} order derivative) of Y . Furthermore we will give almost complete convergence results (with rates) for nonparametric estimates of $f^{x(j)}$. Since $f^x = f^{x(0)}$, we will deduce immediately the convergence of conditional density estimate from the general results concerning $f^{x(j)}$.

The conditional distribution in Ferraty *et al.* (2006) is defined as follows :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))G(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad \forall y \in \mathbb{R}. \quad (4.1)$$

K is a kernel function, G is a distribution function (df) and $h_K := h_{K,n}$ and $h_G = h_{G,n}$ are sequence of positive real numbers which goes to zero as n goes to infinity.

Under a differentiability assumption of $G(\cdot)$, we can obtain the conditional density estimation function as :

$$\widehat{f}^x(y) = \frac{h_G^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (4.2)$$

and its j^{th} partial derivative with respect to y as :

$$\widehat{f}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(j)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (4.3)$$

$G^{(j)}$ is the j^{th} order derivative function of G . We assume that $f^x(\cdot)$ has a unique mode, denoted by $\theta(x)$ which is defined by

$$f^x(\theta(x)) = \sup_{y \in \mathcal{S}} f^x(y). \quad (4.4)$$

A kernel estimator of the conditional mode $\theta(x)$ is defined as the random variable $\hat{\theta}(x)$ which maximizes the kernel estimator $\hat{f}^x(\cdot)$ of $f^x(\cdot)$

$$\hat{f}^x(\hat{\theta}(x)) = \sup_{y \in \mathcal{S}} \hat{f}^x(y). \quad (4.5)$$

Estimating the conditional mode is closely related to the conditional density, and for the last one, the bandwidth selection is very important for the performance of an estimate. The bandwidth must not be too large, so as to prevent over-smoothing, i.e. substantial bias, and must not be too small either, so as prevent detecting the underlying structure. Particularly, in nonparametric curve estimation, the smoothing parameter is crucial for the performance of the estimator.

The k -Nearest Neighbour estimator is a weighted average of response variables in the neighborhood of x , this estimation method take into account the local structure of the data. The k -NN kernel estimate has a significant advantage over the classical kernel estimate. The main advantage of the k -NN method is in the nature of the smoothing parameter. Indeed, in the classical kernel method, the smoothing parameter is the bandwidth h_n , which is a real positive number and in k -NN method, the smoothing parameter k_n takes its values in discrete set.

The existent bibliography of the k -NN method estimation dates back to Royall (1966) and Stone (1977) and has received, since, continuous developments Mack (1981) derived the rates of convergence for the bias and variance as well as asymptotic normality in the multivariate case, Collomb (1980) studied different types of convergence (probability, a.s, a.co) of the estimator of the regression function. Devroye (1978) obtained the strong consistency and the uniform convergence. For the functional data studies, the k -NN kernel estimate was first introduced in the monograph of Ferraty and Vieu (2006), Burba et al. (2009) obtained the rate of almost complete convergence of the regression function using the k -NN method for independent data, Muller and Dippon (2011) obtained the same result in dependent case and the asymptotic normality of robust nonparametric regression function was established in Attouch and Benchikh (2012). For the k -NN conditional mode estimator Attouch and Bouabça (2013) obtained the almost complete convergence with rates in independent and identically distributed (i.i.d.) functional data case.

Now we focus on the estimation of the j^{th} order partial derivative of the conditional density f^x by the k nearest neighbours (k -NN) denoted by $\hat{f}_{kNN}^{x(j)}(y)$ defined as :

$$\hat{f}_{kNN}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))G^{(j+1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(H_{n,k}^{-1}d(x, X_i))}. \quad (4.6)$$

$H_{n,k}(\cdot)$ is defined by :

$$H_{n,k}(x) = \min \left\{ h \in \mathbb{R}^+ / \sum_{i=1}^n \mathbb{I}_{B(x,h_K)}(X_i) = k \right\}.$$

$H_{n,k}$ is a positive random variable (r.v) which depends (X_1, \dots, X_n) and $B(x, h_K) = \{x' \in \mathcal{F} / d(x, x') < h_K\}$.

From now on, when we refer to the bandwidth of the k -NN conditional mode estimation, we mean the number of neighbours k we are considering.

A natural $\widehat{\theta}_{kNN}(x)$ is given by :

$$\widehat{f}^x(\widehat{\theta}_{kNN}) = \sup_{y \in \mathcal{S}} \widehat{f}_{kNN}^x(y). \quad (4.7)$$

Note that this estimate $\widehat{\theta}_{kNN}(x)$ is not necessarily unique, so the remainder of the paper concerns any value $\widehat{\theta}_{kNN}(x)$ satisfying (4.7).

The main goal of this paper is to establish the almost complete convergence for this estimate under α -mixing condition when the nonparametric estimate $\widehat{\theta}_{kNN}(x)$ when the explanatory variable X is valued in infinite dimensional semi-metric vector space \mathcal{F} .

The paper is organized as follows : the following section present the k -NN model estimation of the conditional mode. Then, we give hypotheses and state our main result in Section 3. In Section 4 we illustrate the effectiveness of the k -NN method estimation for real data. All proofs are given in the appendix.

4.2 Notation and assumptions

Definition 1. *The sequence (X_i, Y_i) is said to be α -mixing if*

$$\alpha(n) = \sup_k \sup_{A \in \mathcal{A}_1^k} \sup_{B \in \mathcal{A}_{k+n}^\infty} \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\} \rightarrow 0$$

as $n \rightarrow 0$, where \mathcal{A}_l^m is σ -algebra generated by $\{(X_i, Y_i), l \leq i \leq m\}$.

Now we outline the assumptions that are needed to state our results.

(H1) There exists a nonnegative differentiable function ϕ_x strictly increasing such that :

$$\phi_x(h_K) = \mathbb{P}(X \in B(x, h_K)) > 0.$$

(H2) $\forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$, such that, $|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2})$.
 $b_1, b_2 > 0$ are a strictly positive constant.

(H3) $\exists \xi > 0, f^x \nearrow$ on $(\theta(x) - \xi, \theta(x))$ and $f^x \searrow$ on $(\theta(x), \theta(x) + \xi)$.

(H4) f^x is j -times continuously differentiable in some neighbourhood of y such that

$$\left\{ \begin{array}{l} f^{x^{(l)}}(\theta(x)) = 0, \quad \forall l = 1, \dots, j-1, \\ \text{and} \\ f^{x^{(j)}}(\theta(x)) \neq 0. \end{array} \right.$$

(H5) K is a kernel function with support $(0, 1)$ such that $0 < C_1 < K(t) < C_2 < \infty$.

(H6) G is $(j+1)$ -times continuously differentiable and

$$\left\{ \begin{array}{l} \forall (y_1, y_2) \in \mathbb{R}^2, |G^{(j+1)}(y_1) - G^{(j+1)}(y_2)| \leq C|(y_1) - (y_2)|; \\ \exists \nu > 0, \forall j' \leq j + 1, \lim_{y \rightarrow \infty} |y|^{1+\nu} |G^{(j'+1)}(y)| = 0, \\ G^{(j+1)} \text{ is bounded.} \end{array} \right.$$

(H7) Assume for the sequence of bandwidths $k = k_n$ that there exists a $\gamma \in (0.5, 1)$ such that $k \sim n^\gamma$ and $\frac{\log(n)}{k h_G^{2j+1}} \rightarrow 0$ as $n \rightarrow \infty$.

(H8) $(X_i, Y_i)_{i \in \mathbb{N}}$ is arithmetic α -mixing, namely we have for some

$$C > 0 \text{ and rate } b > 1, \forall n \in \mathbb{N} \quad \alpha(n) \leq Cn^{-b}.$$

(H9) Note by

$$s_{n,1}(x) := \sum_{i,l=1}^n |Cov(\Delta_i(x), \Delta_l(x))| \text{ and } s_{n,2}(x) := \sum_{i,l=1}^n |Cov(\Omega_i \Delta_i(x), \Omega_l \Delta_l(x))|,$$

the terms of covariance, where $\Delta_i(x) = \frac{K(h_K^{-1}d(x, X_i))}{\mathbb{E}[K(h_K^{-1}d(x, X_1))]}$ and $\Omega_i(y) = h_G^{-1}G^{(1)}(h_G^{-1}(y - Y_i))$.

We can split for example $s_{n,2}(x)$ as

$$s_{n,2}(x) := \underbrace{\sum_{i=1}^n Var[(\Omega_i(y) \Delta_i(x))]}_I + \underbrace{\sum_{i,l=1, l \neq i}^n |Cov(\Omega_i(y) \Delta_i(x), \Omega_l(y) \Delta_l(x))|}_{II}. \quad (4.8)$$

Assume for the covariance term $s_n(x) = \max(s_{n,1}(x), s_{n,2}(x))$ that there

$\exists \varrho > 2$, $s_n^{-(b+1)}(x) = o(n^{-\varrho})$, where b is the rate of the mixing coefficient (see hypothesis (H8)).

(H10) This condition is on the distribution of two distinct pairs (X_i, Y_i) and (X_l, Y_l) . We assume that

$$\forall i \neq l : \mathbb{E}[Y_i Y_l | X_i X_l] \leq C < \infty,$$

and the joint distribution functions $\mathbb{P}(X_i \in B(x, h_K), X_l \in B(x, h_K))$ satisfy

$$\exists \epsilon_1 \in (0, 1) : 0 < Q_x(h_K) = O(\phi_x(h_K)^{1+\epsilon_1}),$$

where

$$Q_x(h_K) := \max_{i,l \in \{1, \dots, n\}, i \neq l} \mathbb{P}(X_i \in B(x, h_K), X_l \in B(x, h_K)).$$

(H11) Define $\mathcal{X}(x, h_K) = \max\{1, Q_x(h_K)/(\phi_x(h_K))^2\}$ and assume that

$$\left\{ \begin{array}{l} \lim_{n \rightarrow \infty} h_K = 0, \quad \lim_{n \rightarrow \infty} \frac{\log n}{n h_G \mathcal{X}(x, h_K)} = 0, \\ \text{and} \\ \exists \beta_2 > 0, c_1 > 0, c_2 > 0, c_1 n^{\frac{3-b}{b+1} + \beta_2} \leq \mathcal{X}(x, h_K) \leq c_2 n^{\frac{1}{1-b}}. \end{array} \right.$$

(H12) Define $e = 1/(b + 1)$ where b is the rate of mixing coefficient. Then Assume that

$$\frac{\log(n)\mathcal{X}(x, h_K)^{1-e}n^{1+e}}{h_G k^2} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

Comments on the hypotheses

Hypothesis (H7) is not really more restrictive than in the independent case. However, for their consistency result Burba et al. (2009) need the following two conditions, $k/n \rightarrow 0$ and $\log(n)/k \rightarrow 0$ as $n \rightarrow \infty$; so k must exceed logarithmic order. As in Lian (2011) comments, in most cases in the functional context the small ball probability is of exponential-type. Hence the convergence speed is logarithmic, no matter if the number of neighbours k increases logarithmically or polynomially. For example, if we have for the small ball probability

$$\phi_x(h) \sim \exp\left(-\frac{1}{h^\tau}\right); \text{ then } \phi^{-1}\left(\frac{k}{n}\right) \sim \left(\frac{1}{\log\left(\frac{k}{n}\right)}\right)^\tau$$

where $\phi_x^{-1}(y) = \inf\{h|\phi_x(h) \geq y\}$ (see Lian (2011)). It can be easily seen that the order of k is less important for such small probabilities.

Condition (H10) given in Ferraty and Vieu (2006, p. 163) is not restrictive. For example, if we choose $\mathcal{F} = \mathbb{R}^p$, then $\epsilon_1 = 1$ as soon as each pair of random variables $(X_i; X_l)$ has a bounded density $f_{i;l}$ with respect to the Lebesgue measure.

4.3 Almost complete convergence and almost complete convergence rate

Before presenting the consistency result of the k -NN estimate, the almost complete convergence¹ and the almost complete convergence rate² of $\widehat{\theta}(x)$ (see Ferraty et al. (2006)) is given. All results presented here assumes the α -mixing condition of random variables (X_i, Y_i) (see condition (H8)). We consider two types of non-parametric models. The first one is called continuity-type, which means that the density function f^x is continuous and the second one, the Hölder-type, which means that the density function f^x is Hölder continuous with constants $b_1, b_2 > 0$.

Theorem 4.3.1 (Ferraty and Vieu (2006, p.179)). *In the case of a continuity -type model, we assume condition (H1) for the small ball probability, (H5) for the kernel function, (H7) for the bandwidth k , further more assume hypotheses (H2)- (H3)- (H4), (H6) and (H9). Then we have for $x \in \mathcal{F}$*

$$\lim_{n \rightarrow \infty} \widehat{\theta}(x) = \theta(x). \quad a.co.$$

-
1. We recall that $X_n \rightarrow X$ almost completely if and only if $\forall \epsilon > 0 \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \epsilon) < \infty$.
 2. We recall that $X_n - X = O_{a.co}$ if and only if $\forall \epsilon_0 > 0 \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \epsilon_0 u_n) < \infty$.

Theorem 4.3.1 follows directly from Lemma of Ferraty and Vieu (2006,179) defined as

Lemma 4.3.2. *Under the same conditions of Theorem 4.3.1 we have :*

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathcal{S}} |\widehat{f}^x(y) - f^x(y)| = 0 \quad a.co.$$

Theorem 4.3.3 (Ferraty and Vieu (2006,179)). *Under the same condition of Theorem 4.3.1 and Hölder-type model. Then we have for $x \in \mathcal{F}$*

$$\widehat{\theta}(x) - \theta(x) = O\left(h_K^{b_1} + h_G^{b_2}\right)^{\frac{1}{j}} + O\left(\frac{s_n^2(x) \log(n)}{n^2 h_G}\right)^{\frac{1}{2j}} \quad a.co.$$

Theorem 4.3.3 follows directly by applying Lemma of Ferraty and Vieu (2006,181) defined with

Lemma 4.3.4. *Under the conditions of Theorem 4.3.3 we have :*

$$\sup_{y \in \mathcal{S}} |\widehat{f}^x(y) - f^x(y)| = O\left(h_K^{b_1} + h_G^{b_2}\right) + O\left(\frac{s_{n,2}^2(x) \log(n)}{n^2}\right)^{\frac{1}{2}} \quad a.co.$$

Now we state the almost complete convergence, and almost complete convergence rate result for the nonparametric k -NN conditional mode introduced in (4.7).

Theorem 4.3.5. *In the case of a continuity-type, we suppose that conditions (H1) for the small ball probability, (H5) for the kernel function and (H7) for the bandwidth k , holds.*

If condition (H10) holds with $b > \max\{3/(2\gamma - 1), \epsilon_1 + 2(\gamma - 2)/(\epsilon_1 + 2\gamma - 3)\}$, with $\epsilon_1 > (4 - 2\gamma)$, where γ and ϵ_1 are constants described hereinabove, or assume that condition (H12) is enforced with (H11) and rate $b > \frac{3}{2\gamma - 1}$. Then we have for $x \in \mathcal{F}$:

$$\lim_{n \rightarrow \infty} \widehat{\theta}_{kNN}(x) = \theta(x) \quad a.co. \quad (4.9)$$

Proof

The conditional density $f^x(\cdot)$ is continuous see (H2)-(H3) we get

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y \in \mathcal{S}, |f^x(y) - f^x(\theta(x))| \leq \delta(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon.$$

This leads directly to

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \mathbb{P}(|\widehat{\theta}_{kNN}(x) - \theta(x)| > \epsilon) \leq \mathbb{P}\left(|f^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| \geq \delta(\epsilon)\right). \quad (4.10)$$

By the definition of $\theta(x)$ and $\widehat{\theta}_{kNN}(x)$ we have

$$\begin{aligned}
 |f^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| &= |f^x(\widehat{\theta}_{kNN}(x)) - \widehat{f}^x(\widehat{\theta}_{kNN}(x)) + \widehat{f}^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| \\
 &\leq |f^x(\widehat{\theta}_{kNN}(x)) - \widehat{f}^x(\widehat{\theta}_{kNN}(x))| + |\widehat{f}^x(\widehat{\theta}_{kNN}(x)) - f^x(\theta(x))| \\
 &\leq 2 \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)|.
 \end{aligned} \tag{4.11}$$

It suffices now to combine (4.10) together with (4.11) and Lemma 4.3.6 below, to complete the proof of Theorem 4.3.5 and get the claimed result. ■

Lemma 4.3.6. *Under the conditions of Theorem 4.3.5 we have :*

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| = 0 \quad a.co.$$

Proof

The proof of Lemma 4.3.6 is postponed in appendix. ■

Theorem 4.3.7. *In the case of a Hölder-type model and the same conditions of Theorem 4.3.5 then if (H12) holds instead of (H10) with (H11) and $b > \frac{3}{2\gamma-1}$. Then we have for $x \in \mathcal{F}$*

$$\begin{aligned}
 \widehat{\theta}_{kNN}(x) - \theta(x) &= O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} + h_G^{b_2} \right)^{\frac{1}{j}} + O \left(\frac{\log(n)}{h_G k} \right)^{\frac{1}{2j}} \\
 &+ O \left(\frac{n^{1+e} \log(n)}{h_G k^2} \mathcal{X} \left(x, \phi_x^{-1} \left(\frac{k}{n} \right) \right)^{1-e} \right)^{\frac{1}{2j}} \quad a.co.
 \end{aligned} \tag{4.12}$$

Proof

We have already shown in (4.11) that

$$|f^x(\widehat{\theta}_{kNN}) - f^x(\theta)| = 2 \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)|. \tag{4.13}$$

Writing a Taylor expansion of order j of the function f^x at point $\theta(x)$ and using a first part of (H4), leads to the existence of some θ^* between $\theta(x)$ and $\widehat{\theta}_{kNN}(x)$ such that

$$f^x(\widehat{\theta}_{kNN}(x)) = f^x(\theta(x)) + \frac{1}{j!} f^{x(j)}(\theta^*) (\theta(x) - \widehat{\theta}_{kNN}(x))^j,$$

this result, combined with (4.13), allows one to write that :

$$f^{x(j)}(\theta^*) (\theta(x) - \widehat{\theta}_{kNN}(x))^j = O \left(\sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| \right). \tag{4.14}$$

The continuity of the function $f^{x(j)}$ can be written as :

$$\forall \epsilon > 0, \exists \delta_\epsilon > 0, |\theta(x) - \theta^*| \leq \delta_\epsilon \Rightarrow |f^{x(j)}(\theta(x)) - f^{x(j)}(\theta^*)| \leq \epsilon,$$

this leads directly to

$$\forall \epsilon > 0, \exists \delta_\epsilon > 0, \mathbb{P} \left(|f^{x(j)}(\theta(x)) - f^{x(j)}(\theta^*)| \geq \epsilon \right) \leq \mathbb{P}(|\theta(x) - \theta^*| > \delta_\epsilon).$$

Because the consistency result provided by Theorem 4.3.5 insures that θ^* tends almost completely to $\theta(x)$, therefore :

$$\lim_{n \rightarrow \infty} f^{x(j)}(\theta^*) = f^{x(j)}(\theta(x)) \neq 0. \quad (4.15)$$

By using (4.14), (4.15) together with proposition in Ferraty and Vieu (2006, p.232), we obtain

$$|\theta(x) - \widehat{\theta}_{kNN}(x)|^j = O \left(\sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| \right) \quad a.co, \quad (4.16)$$

and combining 4.16 with Lemma 4.3.8 below complete the proof of Theorem 4.3.7. ■

Lemma 4.3.8. *Under the conditions of Theorem 4.3.7 we have :*

$$\begin{aligned} \sup_{y \in \mathcal{S}} |\widehat{f}_{kNN}^x(y) - f^x(y)| &= O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} + h_G^{b_2} \right)^{\frac{1}{j}} + O \left(\frac{\log(n)}{h_G k} \right)^{\frac{1}{2j}} \\ &+ O \left(\frac{n^{1+e} \log(n)}{h_G k^2} \mathcal{X} \left(x, \phi_x^{-1} \left(\frac{k}{n} \right) \right)^{1-e} \right)^{\frac{1}{2j}} \quad a.co. \end{aligned} \quad (4.17)$$

Proof

The proof of Lemma 4.3.8 is postponed in appendix. ■

The condition (H10) and the condition on the rate b implies that term II in (4.8) decays faster than I.

We get $s_{n,2}^2(x) = O(n(h_G \phi_x(h_K))^{-1})$, see Ferraty and Vieu (2006, p.171). If condition (H11) instead of (H10) is assumed we get three terms for the rate see (4.17). The first one in (4.17) has its origin in the regularity of the conditional mode, the second one stems from term I in (4.8) and the third one represents the dependence of the random variables (compare term II in (4.8))

4.4 Real data application

Now we apply the described method to El Niño time series which gives monthly Sea Surface temperatures. More information and other data-sets can be found about the phenomenon called El Niño in

web site <http://www.cpc.ncep.noaa.gov/data/indices> (see also http://kdd.ics.uci.edu/databases/el_nino/el_nino.data.html). Our study concerns the monthly time series of the Sea Surface Temperature (SST) from June, 1950 up to May, 2004.

A useful way to display such a time series consists in cutting it into 54 pieces or 54 "annual curves" (see Fig. 1). More precisely, let $(X(m), m = 1, \dots, 648)$ be our El Niño time series, m denote the month's number. We can build, for $i = 1, \dots, 54$, the following subsequences : $\forall t \in \{1, \dots, 12\}$, $X_i(t) = X(12 \times (i - 1) + t)$, $X_i = (X_i(1), \dots, X_i(12))$ corresponding to the variations of the SST at the i^{th} year.

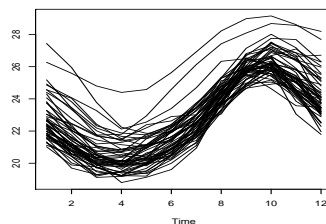


FIGURE 4.1 – El Niño monthly Sea Surface Temperature displayed Year by Year.

However, as described in the real data application application (see Ferraty *et al.* (2006)), the prediction problem can be studied by using the conditional mode approach. Starting from this idea, our objective is to give a comparative study to estimate the conditional mode of the both estimation methods : the kernel method defined in (4.5) and k -NN method defined in (4.7).

For the both methods, we use a quadratic kernel function K , defined by :

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{(0,1)}$$

and the distribution function $G(\cdot)$ is defined by

$$G(x) = \int_{-\infty}^x \frac{15}{4}t^2(1 - t^2)\mathbb{I}_{[-1,1]}(t)dt.$$

We used the semi-metric defined by the L_2 distance between the second derivative of the curves (see Ferraty and Vieu (2002) for more motivations of this choice). We proceed by the following algorithm :

- *Step 1.* We split our data into two subsets :
 - $(X_j, Y_j)_{j=1, \dots, 50}$ training sample,
 - $(X_i, Y_i)_{i=51, \dots, 54}$ test sample.
- *Step 2.* We compute the mode estimator $\hat{\theta}_{X_j}$, for all j by using the training sample.
- *Step 3.* For each X_i in the test sample, we set : $i_* = \arg \min_{j=1, \dots, 50} d(X_i, X_j)$.

– *Step 4.* For all $i = 51, \dots, 54$. We take

$$\widehat{Y}_i = \widehat{\theta}(X_{i^*}),$$

where X_{i^*} is the nearest curve to X_i in the training sample.

And

$$\widetilde{Y}_i = \widehat{\theta}_{k_{opt}}(X_{k_{opt}}),$$

whereas the optimal number of neighbours k_{opt} is defined by

$$k_{opt} = \arg \min_k CV(k),$$

where

$$CV(k) = \sum_{i=51}^{54} (Y_i - \widehat{\theta}_{k_{NN}}^{-i}(x))^2,$$

with

$$\widehat{\theta}_{k_{NN}}^{-i}(x) = \arg \max_y (\widehat{f}_{k_{NN}}^x)^{-i}(y),$$

and

$$(\widehat{f}_{k_{NN}}^x)^{-i}(y) = \frac{h_G^{-1} \sum_{l=1, l \neq i}^{50} K(H_{n,k}^{-1}(x)d(x, X_l))G^{(1)}(h_G^{-1}(y - Y_l))}{\sum_{l=1, l \neq i}^{50} K(H_{n,k}^{-1}(x)d(x, X_l))}.$$

The error used to evaluate this comparison is the mean square error (MSE) expressed by

$$\frac{1}{4} \sum_{i=51}^{54} |Y_i - \widehat{T}(X_i)|^2,$$

where \widehat{T} designate the estimator used : kernel or k -NN method to estimate the conditional mode. The MSE of k -NN mode estimation is 0.00599 and the MSE of the classical mode estimation used in Ferraty et al. (2006) is 0.2664, this result exhibe the effectiveness of the k -NN method. This result will be better if data are heterogeneous as discussed in Burba *et al.* (2009). **Acknowledgement** The authors are grateful to the Editors and Referee for helpful comments and suggestions that improved the paper.

4.5 Appendix

Technical tools

Because the randomness of the smoothing parameter $H_{n,k}$, it is not possible to use the same tools to provide the consistency results as in classical mode estimation. The necessary tools are presented in this section. The following Lemma of Burba et al. (2009) is generalizations of a result firstly presented by Collomb (1980). In our opinion, Lemma 4.5.1 is valid for dependent random variables

as in the original Lemma from Collomb (1980). To overcome this problem, the idea is to frame sensibly $H_{n,k}(x)$ by two non-random windows.

Let $(A_i, B_i)_{i=1,n}$ be a sequence of random variables, not necessarily identically distributed or independent, with values in $(\omega \times \mathbb{R}, \mathcal{A} \otimes \mathcal{B})$.

We note $W : \mathbb{R} \times \omega \rightarrow \mathbb{R}^+$ a measurable function such that :

$$\forall z \in \omega, \forall t, t' \in \mathbb{R} : t \leq t' \Rightarrow W(t, z) \leq W(t', z).$$

Let \mathcal{P} be a real-valued random variable. Then define

$$C_n(\mathcal{P}) = \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n W(\mathcal{P}, A_i) \Omega_i(y)}{\sum_{i=1}^n W(\mathcal{P}, A_i)}.$$

Lemma 4.5.1 (Burba et al.(2009)). *Let (D_n) a sequence of real random variable (r.r.v) and let (v_n) a positive decreasing sequence.*

1. *if $l_i = \lim v_n \neq 0$ and if for every increasing sequence $\beta_n \in]0, 1[$, there exists two sequences of r.r.v $(D_n^-(\beta_n))$ and $(D_n^+(\beta_n))$ such that :*

$$(L'_1) \quad \forall n \in \mathbb{N} : D_n^- \leq D_n^+, \quad \text{and} \quad \mathbb{I}_{\{D_n^- \leq D_n \leq D_n^+\}} \xrightarrow[n \rightarrow +\infty]{a.co.} 1,$$

$$(L'_2) \quad \frac{\sum_{i=1}^n W(D_n^-, A_i)}{\sum_{i=1}^n W(D_n^+, A_i)} - \beta_n = o_{a.co.}(v_n),$$

- (L'_3) *Assume there exists a real positive number c such that*

$$C_n(D_n^-) - c = o_{a.co.}(v_n) \quad \text{and} \quad C_n(D_n^+) - c = o_{a.co.}(v_n).$$

$$\text{Then } C_n(D_n) - c = o_{a.co.}(v_n)$$

2. *If $l_i = 0$ and if for every increasing sequences $\beta_n \in]0, 1[$ with limits 1, (L'_1) , (L'_2) and (L'_3) are verified, the same conclusion holds.*

Burba *et al.* (2009) use the Chernoff-type exponential inequality to check condition (L'_1) in their consistency proof for independent data a. In the α -mixing random variables case, we use the following Lemma of Bradley (see Bosq (1998)) and Lemma4.5.3.

Lemma 4.5.2 (Bosq (1998, p.20)). *Let (X, Y) be a $\mathbb{R}^r \times \mathbb{R}$ valued random vector, such that $Y \in L_p(P)$ for some $p \in [1, \infty]$. Let d a real number such that $\|Y + d\|_p > 0$ and $\epsilon \in (0, \|Y + d\|_p]$. Then there exists a random variable Z such that*

- $\mathbb{P}_Z = \mathbb{P}_Y$ and Z is independent of X ,
- $(\mathbb{P}|Z - Y| > \epsilon) \leq 11 \left(\frac{\|Y + d\|_p}{\epsilon} \right)^{\frac{p}{2p+1}} [\alpha(\sigma(X), \sigma(Y))]^{\frac{p}{2p+1}}$, where $\sigma(X)$ is the σ -algebra generated by X .

The following Lemma is needed in our proofs for technical reasons.

Lemma 4.5.3. *Let (X_i) be an arithmetically α -mixing sequence in the semi-metric space (\mathcal{F}, d) , $\alpha(n) \leq cn^{-b}$, with $b, c > 0$. Define $\Delta_i(x) = \mathbb{I}_{B(x, h_K)}(X_i)$. Then we have*

$$\sum_{i=1}^n \sum_{l=1}^n |\text{Cov}(\Delta_i(x), \Delta_l(x))| = O(n\phi_x(h_K)) + O(\mathcal{X}(x, h_K)^{1-e} n^{1+e});$$

where $\mathcal{X}(x, h_K) = \max\{1, Q_x(h_K)/(\phi_x(h_K))^2\}$ and $e = 1/(b+1)$.

Proof

The proof of this Lemma is similar to that Lemma of Ferraty and Vieu (2004, p.151) except for the choice of the parameter e .

■

Proof of results

Proof [Lemma 4.3.6]

For the demonstration of this result we use Lemma 4.5.1. The main difference to the proof of the independent case in Burba *et al.* (2009) concerns verification of (L'_1) . We need only a small modifications in to verify (L'_2) and (L'_3) . Let

- $v_n = 1$,
- $\forall i = 1, \dots, n : (A_i, B_i) = (X_i, Y_i)$,
- $h_k = D_n = H_{n,k}$,
- $C_n(D_n) = \sup_{y \in \mathcal{S}} \hat{f}_{kNN}^x(y)$, and $c = f^x(y)$.

For $\beta_n \in]0, 1[$, we choose D_n^- et D_n^+ such that :

$$\phi_x(D_n^-) = \sqrt{\beta_n} \frac{k}{n} \quad \text{and} \quad \phi_x(D_n^+) = \frac{1}{\sqrt{\beta_n}} \frac{k}{n}.$$

Verification of (L'_3) :

We put :

$$D_n^- = \phi_x^{-1} \left(\sqrt{\beta_n} \frac{k}{n} \right) \quad \text{and} \quad D_n^+ = \phi_x^{-1} \left(\frac{1}{\sqrt{\beta_n}} \frac{k}{n} \right). \quad (4.18)$$

In the case of $(H11)$, we have

$$s_{n,2}^2(x) = O \left(\frac{n^2}{h_G k} \right) + O \left(h_G^{-1} \mathcal{X}(x, D_n^+)^{1-e} n^{1+e} \right). \quad (4.19)$$

The same is true for D_n^+ . The, we are in position to apply Lemma 4.3.2 to give :

$$\left\{ \begin{array}{l} \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) \Omega_i(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)} \xrightarrow{\text{a.co.}} f^x(y), \\ \sup_{y \in \mathcal{S}} \frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) \Omega_i(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} \xrightarrow{\text{a.co.}} f^x(y). \end{array} \right.$$

Thus condition (L'_3) is verified.

Verification of (L'_2) :

Ferraty and Vieu (2006, p. 162) proved under the conditions of Theorem 4.3.1 that

$$\frac{1}{n \phi_x(h_K)} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h_K} \right) \xrightarrow{\text{a.co.}} 1. \quad (4.20)$$

By (4.20) we have

$$\begin{aligned} \frac{1}{n \phi_x(D_n^-)} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) &\rightarrow 1 \quad \text{a.co.}, \\ \frac{1}{n \phi_x(D_n^+)} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) &\rightarrow 1 \quad \text{a.co.} \end{aligned}$$

Then we obtain :

$$\frac{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} \rightarrow \beta_n.$$

Condition (L'_2) is also verified.

Verification of (L'_1) :

The first part is obvious and for the second one we will show that

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}[|\mathbb{I}_{\{D_n^- < H_{n,k} < D_n^+\}} - 1| > \epsilon] < \infty. \quad (4.21)$$

Let $\epsilon > 0$ be fixed. We know that

$$\mathbb{P}[|\mathbb{I}_{\{D_n^- < H_{n,k} < D_n^+\}} - 1| > \epsilon] \leq \underbrace{\mathbb{P}(H_{n,k} < D_n^-)}_{A_1} + \underbrace{\mathbb{P}(H_{n,k} > D_n^+)}_{A_2}. \quad (4.22)$$

For the two terms A_1 and A_2 in (4.22) we can obtain

$$\begin{aligned}
 A_1 &\leq \mathbb{P} \left[\sum_{i=1}^n \mathbb{I}_{B(x, D_n^-)}(X_i) > k \right] \\
 &\leq \mathbb{P} \left[\left(\sum_{i=1}^n \mathbb{I}_{B(x, D_n^-)}(X_i) - \phi_x(D_n^-) \right) > k - n\phi_x(D_n^-) \right] \\
 &= P_{1n},
 \end{aligned} \tag{4.23}$$

and

$$\begin{aligned}
 A_2 &\leq \mathbb{P} \left[\sum_{i=1}^n \mathbb{I}_{B(x, D_n^-)}(X_i) < k \right] \\
 &\leq \mathbb{P} \left[\left(\sum_{i=1}^n \mathbb{I}_{B(x, D_n^-)}(X_i) - \phi_x(D_n^-) \right) < k - n\phi_x(D_n^-) \right] \\
 &= P_{2n}.
 \end{aligned} \tag{4.24}$$

In the second step of (4.23) and (4.24), we centred the random variables.

At this step, Burba et al. (2009) use the independence of the random variables. The plan here is to split the data into block scheme see Oliveira(2005) or Lu and Cheng (1998). Afterwards we are applying Lemma 4.5.2. Divide the set $\{1, \dots, n\}$ into blocks of length $2J_n$, $m_n = \lceil n/2J_n \rceil$, where $\lceil \cdot \rceil$ is the Gaussian bracket³ and $T_n = n - 2J_n m_n < 2J_n$.

Choosing $m_n \rightarrow \infty$, $T_n \rightarrow \infty$ and J_n is specified later on this proof, see (4.28). By this choice we have $n = 2J_n m_n + T_n$.

The first step is to examine P_{1n} . Let

$$U_n(l) = \left(\sum_{i=(l-1)J_n+1}^{lJ_n} \mathbb{I}_{B(x, D_n^-)}(X_i) - \phi_x(D_n^-) \right).$$

Let us define

$$B_{n1} = \sum_{l=1}^{m_n} U_n(2l-1), \quad B_{n2} = \sum_{l=1}^{m_n} U_n(2l) \quad \text{and} \quad R_n = \left(\sum_{i=2J_n m_n+1}^n \mathbb{I}_{B(x, D_n^-)}(X_i) - \phi_x(D_n^-) \right).$$

We obtain

$$\begin{aligned}
 P_{1n} &\leq \mathbb{P} \left(B_{n1} > \frac{k-n\phi_x(D_n^-)}{3} \right) + \mathbb{P} \left(B_{n2} > \frac{k-n\phi_x(D_n^-)}{3} \right) + \mathbb{P} \left(R_n > \frac{k-n\phi_x(D_n^-)}{3} \right) \\
 &= P_{1n}^{(1)} + P_{1n}^{(2)} + P_{1n}^{(3)}.
 \end{aligned} \tag{4.25}$$

Regarding $P_{1n}^{(1)}$, using Lemma 4.5.2 with $d = J_n m_n$ leads to

$$0 < J_n m_n \leq \|U_n(2l-1) + d_n\|_\infty \leq 2J_n + J_n m_n.$$

3. $\lceil x \rceil = \max\{y \in \mathbb{Z}, z \leq x\}; x \in \mathbb{R}$

Because $J_n m_n = O(n)$ and $k/n \rightarrow 0$, we can write

$$\varepsilon = \frac{k - n\phi_x(D_n^-)}{6m_n} = \frac{k(1 - \sqrt{B})}{6m_n} \in (0, \|U_n(2l-1) + d_n\|_\infty].$$

This choice of ε is motivated by (4.27) below. We can construct $\tilde{U}_n(2l-1)_{l=1}^{m_n}$ by Lemma 4.5.2 such that

- the random variables $\tilde{U}_n(2l-1)_{l=1}^{m_n}$ are independent,
- $\tilde{U}_n(2l-1)_{l=1}^{m_n}$ and $U_n(2l-1)_{l=1}^{m_n}$ has the same distribution
- and

$$\mathbb{P}\left(|\tilde{U}_n(2l-1) - U_n(2l-1)| > \varepsilon\right) \leq 11 \left(\frac{\|U_n(2l-1) + d\|_\infty}{\varepsilon}\right)^{\frac{1}{2}} \sup |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|,$$
 where the supremum is taken over all sets

$$A; B \in \sigma(U_n(1), U_n(3), \dots, U_n(2m_n - 1)).$$

This leads to

$$\begin{aligned} P_{1n} &= \mathbb{P}\left(\sum_{l=1}^{m_n} [\tilde{U}_n(2l-1) + (U_n(2l-1) - \tilde{U}_n(2l-1))] > \frac{k - n\phi_x(D_n^-)}{3}\right) \\ &\leq \mathbb{P}\left(\sum_{l=1}^{m_n} \tilde{U}_n(2l-1) > \frac{k - n\phi_x(D_n^-)}{6}\right) + \mathbb{P}\left(\sum_{l=1}^{m_n} (U_n(2l-1) - \tilde{U}_n(2l-1)) > \frac{k - n\phi_x(D_n^-)}{6}\right) \\ &= P_{1n}^{(11)} + P_{1n}^{(12)}. \end{aligned} \quad (4.26)$$

Now we deal with $P_{1n}^{(12)}$. By Lemma 4.5.2, we have

$$P_{1n}^{(12)} \leq \mathbb{P}\left(\sum_{l=1}^{m_n} (U_n(2l-1) - \tilde{U}_n(2l-1)) > \frac{k - n\phi_x(D_n^-)}{6m_n}\right) \leq C \frac{n^2}{J_n^{\frac{3}{2}} k} \alpha(J_n). \quad (4.27)$$

We choose the sequence J_n such that

$$J_n^a = \frac{n^2}{2a_r a k}, \quad (4.28)$$

where r is a positive constant specified below and $a > 2/\gamma - 1$.

By the condition on the mixing coefficient b and simple algebra allows us to have the following result

$$\frac{n^2}{J_n^{\frac{3}{2}} k} \alpha(J_n) = C n^{(2-\gamma)(a-3/2-b)/a}$$

leads directly to

$$\sum_{n=1}^{\infty} P_{1n}^{(12)} < \infty. \quad (4.29)$$

Now we apply the inequality of Markov on term $P_{1n}^{(11)}$ for some $t > 0$,

$$\mathbb{P}\left(\sum_{a=1}^{m_n} \tilde{U}_n(2l-1) > \frac{k - n\phi_x(D_n^-)}{6}\right) \leq \exp\left(-t \frac{k - n\phi_x(D_n^-)}{6}\right) \mathbb{E}\left[\exp\left(t \sum_{l=1}^{m_n} \tilde{U}_n(2l-1)\right)\right]. \quad (4.30)$$

Using the independence of the random variables $(\tilde{U}_n(2l-1))_{l=1}^{m_n}$ we have

$$\mathbb{E} \left[\exp \left(t \sum_{l=1}^{m_n} \tilde{U}_n(2l-1) \right) \right] = \prod_{l=1}^{m_n} \mathbb{E} \left[\exp \left(t \tilde{U}_n(2l-1) \right) \right]. \quad (4.31)$$

Taking $t = r \log n / (h_G k)$ and the same choice of J_n its follows that

$$t |\tilde{U}_n(2l-1)| \leq \frac{2r J_n \log n}{h_G k} = \log n \left(\frac{n^2}{h_G k^{a+1}} \right)^{\frac{1}{a}}. \quad (4.32)$$

Take now $a > 2/\gamma - 1$ and $k \sim n^\gamma$ we obtain for n large enough $t |\tilde{U}_n(2l-1)| \leq 1$ and

$$\exp(t \tilde{U}_n(2l-1)) \leq 1 + t \tilde{U}_n(2l-1) + t^2 \tilde{U}_n(2l-1)^2.$$

The random variable $\tilde{U}_n(2l-1)$ has the same distribution as the centred random variable $U_n(2l-1)$. Hence we know that the expectation of the linear term is zero, $\mathbb{E}[\tilde{U}_n(2l-1)] = 0$ with $\exp(x) \geq 1 + x$ we get

$$\mathbb{E} \left[\exp(t \tilde{U}_n(2a-1)) \right] \leq 1 + \mathbb{E} \left[t^2 \tilde{U}_n(2a-1)^2 \right] \leq \exp \left(t^2 \mathbb{E} \left[\tilde{U}_n(2a-1)^2 \right] \right). \quad (4.33)$$

Furthermore, because $\tilde{U}_n(2l-1)$ and $U_n(2l-1)$ have the same distribution function and by some calculations, it follows that

$$\sum_{l=1}^{m_n} \mathbb{E} \left[\tilde{U}_n(2l-1)^2 \right] \leq \sum_{i;l=1}^n \left| \text{Cov}(\mathbb{I}_{B(x, D_n^-)}(X_i), \mathbb{I}_{B(x, D_n^-)}(X_l)) \right|. \quad (4.34)$$

Since $\phi_x(D_n^-) = \sqrt{\beta_n \frac{k}{n}}$ and $k \sim n^\gamma$ we know that $\phi_x(D_n^-) = O(n^{\gamma-1})$.

We apply Lemma 4.5.3 and get

$$\sum_{l=1}^{m_n} \mathbb{E} \left[\tilde{U}_n(2l-1)^2 \right] \leq C_1 \sqrt{\beta_n} k + C_2 \mathcal{X}(D_n^-)^{1-e} n^{1+e}. \quad (4.35)$$

Below, we present the arguments if condition (H12) holds, because in the case of (H10) the rationale follows as the same line. By (4.31), (4.33), (4.35), and $t = r \log n / (h_G k)$ we have that

$$\mathbb{E} \left[\exp \left(t \sum_{a=1}^{m_n} \tilde{U}_n(2a-1) \right) \right] \leq \exp \left(C_1 \sqrt{\beta_n} r^2 \frac{(\log n)^2}{h_G^2 k} \right) \cdot \exp \left(C_2 \sqrt{\beta_n} r^2 \frac{(\log n)^2 \mathcal{X}(D_n^-)^{1-s} n^{1+s}}{h_G^2 k^2} \right). \quad (4.36)$$

Because $k \sim n^\gamma$, the first term in (4.36) satisfies

$$\exp \left(C_1 \sqrt{\beta_n} r^2 \frac{(\log n)^2}{h_G^2 k} \right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

If (H12) holds, we have for the second term in (4.36)

$$\exp \left(C_2 \sqrt{\beta_n} r^2 \frac{(\log n)^2 \mathcal{X}(D_n^-)^{1-e} n^{1+e}}{h_G^2 k^2} \right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

Since $\phi_x(D_n^-) = \sqrt{\beta_n} \frac{k}{n}$ and by choosing $r > 6/(1 - \sqrt{B})$, the first term in (4.30) becomes

$$\exp\left(-t \frac{k - n\phi_x(D_n^-)}{6}\right) = \exp\left(\frac{-r(1 - \sqrt{B})}{6} \log n\right) = n^{\frac{-r(1 - \sqrt{B})}{6}}. \quad (4.37)$$

So

$$\sum_{n=1}^{\infty} P_{1n}^{(11)} < \infty. \quad (4.38)$$

Now, combine relations (4.29) and (4.38) to obtain

$$\sum_{n=1}^{\infty} P_{1n}^{(1)} \leq \sum_{n=1}^{\infty} P_{1n}^{(11)} + \sum_{n=1}^{\infty} P_{1n}^{(12)} < \infty.$$

By similar arguments as for $P_{1n}^{(1)}$ we deduce that

$$\sum_{n=1}^{\infty} P_{1n}^{(2)} < \infty.$$

The final step is to examine

$$P_{1n}^{(3)} = \mathbb{P}\left(R_n > \frac{k - n\phi_x(D_n^-)}{3}\right).$$

We know that $|R_n| \leq 4J_n$ and $k - n\phi_x(D_n^-)/3 = O(k)$. Together with the choice of J_n in (4.28) and the condition on the parameter $a > 2/\gamma - 1$ we have $k > J_n$ for large n . This implies

$$\sum_{n=1}^{\infty} P_{1n}^{(3)} < \infty.$$

We get Finally

$$\sum_{n=1}^{\infty} P_{1n} \leq \sum_{n=1}^{\infty} P_{1n}^{(1)} + \sum_{n=1}^{\infty} P_{1n}^{(2)} + \sum_{n=1}^{\infty} P_{1n}^{(3)} < \infty.$$

Analysis of P_{2n} is similar to that of P_{1n} . This finishes the proof of condition (L'_1) , which states that

$$\mathbb{P}\left(\left|\mathbb{I}_{\{D_n^- < H_{n,k} < D_n^+\}}\right|\right) \rightarrow 1 \quad \text{almost completely.}$$

Now, we are in the position to apply Lemma 4.5.1 to get the claimed result

$$\lim_{n \rightarrow \infty} \widehat{\theta}_{kNN}(x) = \theta(x) \quad a.co.$$

■

Proof[Lemma 4.3.8]

To prove the Lemma 4.3.8, we will use the Lemma 4.5.1. The verification of (L'_1) and (L'_2) are checked in a similar way as in the proof of Theorem 4.3.5.

For (L'_1) , we use Lemma 4.3.4 with notation (4.18) and (4.19), then we get

$$\left\{ \begin{array}{l} \left| \frac{\sup_{y \in \mathcal{S}} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right) \Omega_i(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^-} \right)} - \sup_{y \in \mathcal{S}} f^x(y) \right| = O((D_n^-)^{b_1}) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right) \\ \left| \frac{\sup_{y \in \mathcal{S}} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right) \Omega_i(y)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{D_n^+} \right)} - \sup_{y \in \mathcal{S}} f^x(y) \right| = O((D_n^+)^{b_1}) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right) \end{array} \right.$$

In the other word :

$$\left\{ \begin{array}{l} \left| C_n(D_n^-) - \sup_{y \in \mathcal{S}} f^x(y) \right| = O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} \right) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right) \\ \left| C_n(D_n^+) - \sup_{y \in \mathcal{S}} f^x(y) \right| = O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} \right) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right) \end{array} \right.$$

So : $\exists c = \sup_{y \in \mathcal{S}} f^x(y) > 0$, such that :

$$\left\{ \begin{array}{l} |C_n(D_n^-) - c| = O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right) \\ |C_n(D_n^+) - c| = O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right) \end{array} \right.$$

So (L'_3) is verified .

We have finally :

$$\sup_{y \in \mathcal{S}} \left| \widehat{f}_{kNN}^x(y) - f^x(y) \right| = O \left(\phi_x^{-1} \left(\frac{k}{n} \right)^{b_1} \right) + O(h_G)^{b_2} + O_{\text{a.co.}} \left(\sqrt{\frac{s_{n,2}^2(x) \log(n)}{n^2}} \right).$$

To check condition (L'_2) we estimate $s_{n,2}^2(x)$ by bounds obtained either by (H12), see (4.19). This completes the proof of this Lemma. \blacksquare

Références

1. Attouch M., Bouabça, W. (2013). The k -nearest neighbors estimation of the conditional mode for functional data. *Rev. Roumaine Math. Pures Appl.*, **58**, No. 4, pp. 393-415.
2. Attouch M., Benchikh, T. (2012). Asymptotic distribution of robust k -nearest neighbour estimator for functional nonparametric models. *Mathematic Vesnic*, **64**, No. 4, pp. 275-285.
3. Bosq, D. (1998). *Nonparametric statics for stochastic processes : estimation and prediction*. Lecture notes in statics, Springer.
4. Burba, F., Ferraty, F., Vieu, P. (2008). Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique. (French) [Convergence of the k nearest neighbor kernel estimator in nonparametric functional regression]. *C. R. Math. Acad. Sci. Paris*, **346**, no. 5-6, 339-342.
5. Burba, F., Ferraty, F., Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression, *J. Nonparametr. Stat.*, **21**(4), 453-469.
6. Cheng, B, Robinson, PM. (1991). Density estimation in strongly dependent non-linear time series. *Statistica Sinica*, 1 :335-359.
7. Collomb, G. (1980). Estimation de la regression par la methode des k plus proches avec noyau : quelques propriétés de convergence ponctuelle. *Statistique Non Parametrique Asymptotique*, J-P. Raoult, Ed. vol.821 of *Lecture Notes in Mathematics*. Springer Berlin/Heidelberg, pp. 159-175.
8. W., Hassani, S. (1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf.*, **15**, 227-236.
9. Dabo-Niang, S., Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C. R., Math., Acad. Sci. Paris*, **344**, 49-52.
10. Devroye, L.P. (1978). The uniform convergence of nearest neighbour regression function estimators and their application in optimization. *IEEE Trans. Inform. Theory*, **24**, 142-151.
11. Estevas, G., Vieu, P. (2003). Nonparametric estimation under long memory dependence. *Journal of Nonparametric Statistics* , 15 :535-551.
12. Ezzahrioui, M., Ould-Saïd, E. (2006). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. *Preprint, LMPA No 277, Univ. du Littoral Côte d'Opale*.
13. Ezzahrioui, M., Ould-Saïd, E. (2010). Somo Asymptotic result of nonparametric estimator of the conditional mode for functional time-series data. *Journal of Statistica Neerlandica*, **64**, 171-201.
14. Ezzahrioui, M., Ould-Saïd, E. (2008). Asymptotic normality of nonparametric estimator of the conditional mode for functional data. *Journal of Nonparametric Statistics*, **20**, 3-18.

15. Ferraty, F., Laksaci, A., Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Processes*, **9**, 47-76.
16. Ferraty, F., Mas, A., Vieu, P. (2007). Nonparametric regression on functional data : Inference and practical aspects. *Australian and New Zealand Journal of Statistics*, **49**, 267-286.
17. Ferraty, F., Vieu, Ph. (2002). The functional nonparametric models and application to spectrometric data. *Compute.Stat.*, **17**, No.12, 1025-1028.
18. Ferraty, F., Vieu, Ph. (2004). Non-parametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of Nonparametric Statistics*, **16**, 111-125.
19. Ferraty, F., Vieu, Ph. (2006). *Nonparametric Functional Data Analysis*. Springer, New York.
20. Giraitis, L., Koul, H. L., Surgailis, D. (1996). Asymptotic normality of regression estimators with long memory errors. *Statistics and Probability Letters*, **29** :317-335.
21. Gasser, T., Hall, P., Presnell, B. (1998). *Nonparametric estimation of the mode of a distribution of random curves*. J. Roy. Statist. Soc., Ser B, **60** :681-691.
22. Hart, J. (1984). Efficiency of a kernel density estimator under an autoregressive dependence model. *Journal of the American Statistical Association*, **83** :86-99.
23. Li, W.V., Shao, Q.M. (2001). *Gaussian processes : inequalities, small ball probabilities and applications*. In Stochastic processes : Theory and Methods, Ed. C.R. Rao and D. Shanbhag. Handbook of Statistics, **19**, North-Holland, Amsterdam.
24. Lian, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electronic Journal Of Statics* **5**,31-40.
25. Lu, Z., Cheng, P. (1998). Strong consistency of nearest neighbor kernel regression estimation for stationary dependent samples. *Science in China Series A : Mathematics*, **41**, **9**, 918-926.
26. Mack, Y. P. (1981). Local properties of k -NN regression estimates. *SIAM J. Algebr Discrete Methods*, **2** , 311-323.
27. Mielniczuk, J. (1997). On the asymptotic mean integrated squares error of kernel density estimator for dependent data. *Statistics and Probability Letters*, **34** :53-58.
28. Müller, S., Dippon, J. (2011). K-NN Kernel Estimate for Nonparametric Functional Regression in Time Series Analysis. *Fachbereich Mathematic*. Universität Stuttgart.
29. Oliveira, P. E. (2005). Nonparametric density and regression estimation functional data. *Tech. rep., Departamento de Matematica*, Universidade de Coimbra.
30. Ould-Saïd, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Statist.*, **24**, 231-239.
31. Robinson, PM. (1983). Nonparametric estimates for time series. *Journal of Time Series Analysis*, **4** :185-201.

32. Rosenblatt, M. (1970). *Density estimates and markov sequences*. In M. Puri, editor, Non-parametric techniques in statistical inference, pages 199-210. Cambridge University Press, London.
33. Royall, R.M. (1966). *A class of nonparametric estimates of a smooth regression function*, Ph.D. Diss., Stanford University.
34. Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist*, **5**, 595-645.

Chapitre 5

Normalité asymptotique : Cas α -mélange

Ce travail traite la normalité asymptotique de l'estimateur du mode conditionnel lorsque les données sont fonctionnelles dépendantes. Cet travail fait l'objet d'un article soumis.

Some asymptotic normality result of k -Nearest Neighbour estimator of the conditional mode function for functional data under dependency

Attouch Mohammed Kadi¹ and Bouabça Wahiba²

^{1,2}Univ. Sidi Bel Abbès, Laboratoire de Statistique Processus Stochastiques, Algérie,
attou_kadi@yahoo.fr, wahiba_bouab@yahoo.fr

Abstract In this paper, we study the nonparametric estimator of the conditional mode using the k -Nearest Neighbors (k-NN) estimation method for a scalar response variable given a random variable taking values in a semi-metric space. We establish the asymptotic normality for functional data under dependency of a predictor, and propose confidence bands for the conditional mode function. Some simulations have been driven to show how our methodology can be implemented.

Key words : Functional data ; k-NN estimator ; the conditional mode estimation ; small balls probability ; asymptotic normality.

5.1 Introduction

Let $(X_i, Y_i), i = 1, \dots, n$ be n copies of random vector identically distributed as (X, Y) where X is valued in infinite dimensional semimetric vector space (\mathcal{F}, d) and Y 's are valued in \mathbb{R} , We are interested in predicting Y from the data for a fixed value of X . We do not suppose the existence of a density for the functional random variable (f.r.v) X .

Such an approach has been investigated by several authors when the observed data are i.i.d. or when the process is mixing (see the surveys by Collomb (1985) and Györfi et al. (1989).

However, we know that if the conditional distribution of Y given X has a dominant center peak and a smaller peak far from the center, then it is more reasonable to consider the conditional mode function.

The estimation of the conditional mode has a long history and has been studied by many authors in the statistics literature, Romano (1998) investigated the asymptotic behavior of the kernel estimate of the conditional mode with data dependent bandwidths and obtained result under weaker smoothness on the underlying density $\hat{F}_D(\cdot)$.

Vieu (1996) obtained a rate of convergence for both local and global estimates of the mode function. Under right random censoring, Louani (1998) studied the asymptotic normality of the kernel estimator of the mode. When conditioning by one of the coordinates of a bidimensional random vector in the i.i.d random vector case, Samanta and Thavaneswaran (1990) showed that under some regularity conditions, the kernel estimator of the conditional mode function is consistent and asymptotically normally distributed. Finally, Mehra et al.(2000) established the law of iterated logarithm, uniform

almost sure convergence over a compact set, and asymptotic normality of the smoothed rank nearest neighbors estimator of the conditional mode function.

For the dependent case, the strong consistency of the conditional mode estimator was established under ϕ -mixing condition by Collomb et al.(1987) and their results can be applied to process forecasting. In the α -mixing case, the strong uniform consistency over a compact set and the asymptotic normality were obtained by Ould-Saïd (1993) and Louani and Ould-Saïd (1999), respectively. In a general ergodic framework, a process prediction via conditional mode estimation was described by Ould-Saïd (1997) and its strong consistency was obtained.

There exists many instances where conditional densities of the regression function becomes zero anywhere and therefore can not be used in some problems implying prediction. Starting from this point of view, Ould Saïd (1997) in finite dimensional spaces and Ferraty et al. (2006) in infinite dimensional spaces show that the conditional mode provides an alternative approach to the classical regression function. The first result of the asymptotic normality on the kernel estimator comes from Masry (2005), he considers the case of α -mixing data , but he did not give the explicit expression of the dominant asymptotic terms of bias and variance. After Ferraty et al.(2007) gives the explicit expression of the asymptotic law (that is the dominant terms of bias and variance) In the case of set of independent data . The results exposed in the papers of Delsol (2008b) make the link between these articles, they generalize the results of Ferraty et al. (2007) in the case of data α -mixing, Attouch and Benchikh (2012) established the asymptotic normality of robust nonparametric regression function.

The k -NN kernel estimate is a widely studied estimate if the explanatory variable is an element of a finite-dimensional space, see Györfi et al (2002). In the functional case with real-valued response, two different approaches for the k -NN regression estimation exist. An important advantage of the k -NN approach is its local adaptation, a property that is not enjoyed by any kernel method for example. For these reasons the k -NN approach has already been used in other semiparametric settings assuming i.i.d. sampling, see e.g. Robinson ((1987), (1995)), Newey (1990), Delgado (1992), Delgado and Stengos (1994), Liu and Lu (1997), Heckman ,Ichimura and Todd (1998), Abadie and Imbens (2006), Pinkse (2006), Li (2006), Jacho-Chávez (2008), and Jun and Pinkse (2009a, 2009b, 2009c). Although results are available for the nonparametric k -NN estimation of densities and regression functions with time series, see e.g. Yakowitz (1987), Boente and Fraiman (1990,1998), Tran and Yakowitz (1993) among many others, and for the k -NN conditional mode estimator Attouch and Bouabça (2013) obtained the almost complete convergence with rates in independent and identically distributed (i.i.d.) functional data case. We have not seen any previous use of the k -NN method in nonparametric estimation for conditional mode with dependent data as in this paper.

Our aim is to build nonparametric estimates of several function related with the conditional probability distribution of Y given X .

For $x \in \mathcal{F}$ we will denote the conditional distribution function of Y by :

$$\forall y \in \mathbb{R} \quad F^x(y) = \mathbb{P}(Y \leq y | X = x).$$

Then we will denote by f^x (resp $f^{x(j)}$) the conditional density (resp its j^{th} order derivative) of Y . Furthermore we will give almost complete convergence results (with rates) for nonparametric estimates of $f^{x(j)}$. Since $f^x = f^{x(0)}$, we will deduce immediately the convergence of conditional density estimate from the general results concerning $f^{x(j)}$.

In the following, x will be a fixed in \mathcal{F} ; C and $C(x)$ are generic constants with $C(x)$ depending on x , N_x will denote a fixed neighborhood of x , \mathcal{S} will be a fixed compact subset of \mathbb{R} , and any real function with an integer in brackets as exponent denotes its derivative with the corresponding order. The conditional distribution in Ferraty et al. (2006) is defined as follows :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))G(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad \forall y \in \mathbb{R}. \quad (5.1)$$

K is a kernel function, G is a distribution function (df) and $h = h_K := h_{K,n}$ and $h_G = h_{G,n}$ are sequence of positive real numbers which goes to zero as n goes to infinity.

Under a differentiability assumption of $G(\cdot)$, we can obtain the conditional density estimation function as :

$$\widehat{f}^x(y) = \frac{h_G^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (5.2)$$

and its j^{th} partial derivative with respect to y as :

$$\widehat{f}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))G^{(j)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (5.3)$$

$G^{(j)}$ is the j^{th} order derivative function of G . We assume that $f^x(\cdot)$ has a unique mode, denoted by $\theta(x)$ which is defined by

$$f^x(\theta(x)) = \sup_{y \in \mathbb{R}} f^x(y). \quad (5.4)$$

A kernel estimator of the conditional mode $\theta(x)$ is defined as the random variable $\widehat{\theta}(x)$ which maximizes the kernel estimator $\widehat{f}^x(\cdot)$ of $f^x(\cdot)$

$$\widehat{f}^x(\widehat{\theta}(x)) = \sup_{y \in \mathbb{R}} \widehat{f}^x(y). \quad (5.5)$$

Let consider quantities

$$\widehat{f}^x(y) = \frac{\widehat{f}_N^x(y, h)}{\widehat{F}_D^x(h)}. \quad (5.6)$$

Where

$$\widehat{f}_N^x(y, h) = \frac{1}{n h_G \phi_x(h)} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)) G^{(1)}(h_G^{-1}(y - Y_i)), \quad (5.7)$$

$$\widehat{F}_D^x(h) = \frac{1}{n \phi_x(h)} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)), \quad (5.8)$$

where $\phi_x(\cdot)$ is a function supposed to be strictly positive and it will be described later.

An analogous estimator to (5.7) was already given in Ferraty et al. (2006) in the general setting, remark that (5.7) and (5.8) are unbiased estimators of $a_1 f^x(y) F_D^x(h) = a_1 f_N^x(y, h)$ and $a_1 F_D^x$ respectively where a_1 and F_D^x will be described later.

Now we focus on the estimation of the j^{th} order derivative of the conditional density f on x by the k nearest neighbors (k -NN), where $\widehat{f}_{kNN}^{x(j)}(y)$ is defined by :

$$\widehat{f}_{kNN}^{x(j)}(y) = \frac{h_G^{-j-1} \sum_{i=1}^n K(H_{n,k}^{-1} d(x, X_i)) G^{(j+1)}(h_G^{-1}(y - Y_i))}{\sum_{i=1}^n K(H_{n,k}^{-1} d(x, X_i))}. \quad (5.9)$$

$H_{n,k}(\cdot)$ is defined by :

$$H_{n,k}(x) = \min \left\{ h \in \mathbb{R}^+ / \sum_{i=1}^n \mathbb{I}_{B(x,h)}(X_i) = k \right\}.$$

$H_{n,k}$ is a positive random variable (r.v) which depends (X_1, \dots, X_n) , and $B(x, h) = \{x' \in \mathcal{F} / d(x, x') < h\}$.

so that

$$\widehat{f}_{kNN}^x(y) = \frac{\widehat{f}_N^x(y, H_{n,k})}{\widehat{F}_D^x(H_{n,k})}. \quad (5.10)$$

A natural $\widehat{\theta}_{kNN}(x)$ is given by :

$$\widehat{f}^x(\widehat{\theta}_{kNN}) = \sup_{y \in \mathbb{R}} \widehat{f}_{kNN}^x(y). \quad (5.11)$$

Note that this estimate $\widehat{\theta}_{kNN}(x)$ is not necessarily unique, so the remainder of the paper concerns any value $\widehat{\theta}_{kNN}(x)$ satisfying (??).

We consider two types of non-parametric models. The first one is called continuity-type, which means that the density function f^x is continuous and the second one, the Hölder-type, which means that the density function f^x is Hölder continuous with constants $b_\gamma > 0$.

Our main goal is to establish the asymptotic normality of the estimator in (??) when suitably normalized. As a consequence we get the asymptotic normality of a predictor and propose confidence bands for the conditional mode function with the k -NN method. The paper is organized as follows : in Section 2 we give the assumptions and main results. After we introduce the asymptotic normality in Section 3, some applications are given in Section 4. Finally, the proof of the main result is relegated to Section 5 with some auxiliary results.

5.2 Notation and assumptions

Definition 2. The sequence (X_i, Y_i) is said to be α -mixing if

$$\alpha(n) = \sup_k \sup_{A \in \mathcal{A}_1^k} \sup_{B \in \mathcal{A}_{k+n}^\infty} \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\} \rightarrow 0 \quad \text{as } n \rightarrow 0, \quad (5.12)$$

where \mathcal{A}_l^m is σ -algebra generated by $\{(X_i, Y_i), l \leq i \leq m\}$.

Furthermore let \mathcal{S} be a compact set of \mathbb{R} such that $\theta(x) \in \overset{\circ}{\mathcal{S}}$, where $\overset{\circ}{\mathcal{S}}$ denotes the interior of \mathcal{S} .

Now we outline the assumptions.

(H1) There exist two functions $\psi_x(\cdot), \zeta_0^x(\cdot)$ such that

1. For all $h > 0$ $\phi_x(h) = \mathbb{P}(W_i \leq h) = \mathbb{P}(X_i \in B(x, h)) > 0$ and $\lim_{h \rightarrow 0} \phi_x(h) = 0$,
2. $\forall u \in [0, 1], \lim_{h \rightarrow 0} \frac{\phi_x(uh)}{\phi_x(h)} = \lim_{h \rightarrow 0} \zeta_h(u) = \zeta_0(u)$.
3. $\sup_{i \neq j} \mathbb{P}[(X_i, Y_i) \in B(x, h) \times B(x, h)] = \mathbb{P}[W_i \leq h, W_j \leq h] \leq \psi_x(h)$,
where $\psi_x(h) \rightarrow 0$ as $h \rightarrow 0$. Also, we assume that $\psi_x(h) = O(\phi_x^2(h))$.

(H2) The conditional density and its two first derivatives satisfy the Hölder condition

$$\forall (y_1, y_2) \in \mathbb{R}^2, \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, |f^{x_1(j)}(y_1) - f^{x_2(j)}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}).$$

for $j = 0, 1, 2$ where \mathcal{N}_x is a neighborhood of x , and C_x is constant that depends on x .

(H3) The mixing coefficient satisfies

$$\sum_{k=1}^{\infty} k^\delta (\alpha(k))^{1-(2/\tau)} < +\infty \quad \text{for some } \tau > 2 \text{ and } \delta > 1 - \frac{2}{\tau}.$$

(H4) The kernel K is nonnegative, with compact support $[0, 1]$ of class C^1 on $[0, 1]$, $K(0) > 0$ and $K(1) > 0$ and its derivative K' exists on $[0, 1]$ and $K'(t) < 0$.

(H5) $G^1(\cdot)$ is twice differentiable such that

1. $\forall (y_1, y_2) \in \mathbb{R}^2, |G^{(j)}(y_1) - G^{(j)}(y_2)| \leq C|y_1 - y_2|$ for $j = 1, 3$
2. $\int |G^{(j)}(u)|^k du < +\infty$ for $j = 1, 3, k = 1$ and for $j = 2, k = 1, 2, \tau$.
3. $G^{(j)}$ is bounded for $j = 1, 2, 3$. Furthermore, we assume that $m = \inf_{[0,1]} K(t)G^{3(t)} \neq 0$

(H6) The bandwidth h_G satisfies, for $0 \leq j \leq 2$

1. $h_G \rightarrow 0$ $nh_G^{2j+1}\phi_x^2(h) \rightarrow +\infty$ and $\frac{\log^2(n)}{nh_G^{j+2}\phi_x(h)} \rightarrow 0$, as $n \rightarrow \infty$.
2. $n^\zeta h_G^{j+2} \rightarrow +\infty$ as $n \rightarrow \infty$ for some $\zeta > 0$,
3. $nh_G^{2j+2}\phi_x^2(h) \rightarrow 0$ as $n \rightarrow \infty$

(H7) There exist sequences of integers (u_n) and (v_n) increasing to infinity such that $(u_n + v_n) \leq n$, satisfying

1. $v_n = (o(nh_G\phi_x(h))^{1/2})$ and $\left(\frac{n}{h_G\phi_x(h)}\right)^{1/2} \alpha(v_n) \rightarrow 0$ as $n \rightarrow \infty$,

2. $q_n v_n = (o(nh\phi_x(h))^{1/2})$ and $q_n \left(\frac{n}{h\phi_x(h)} \right)^{1/2} \alpha(v_n) \rightarrow 0$ as $n \rightarrow \infty$, where q_n is the largest integer such $q_n(u_n + v_n) \leq n$

(H8) The sequence of positive real numbers $k_n = k$ satisfies :

$$\frac{k}{n} \rightarrow 0 \text{ and } \frac{\log(n)}{k} \rightarrow 0, \frac{\log(n)}{k h_G^{2j+1}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Comments on the hypotheses :

1. The assumption (H1) is an assumption commonly used for the explanatory variable X (see Ferraty and Vieu (2006) for more details and examples).
2. Assumption (H2) characterize the structural functional space of our model and is needed to evaluate the bias term in our asymptotic properties.
3. Assumptions (H3), and (H6) are classical in the functional estimation in finite or infinite dimension spaces. The choice of the sequences (u_n) , (v_n) and (q_n) in (H7) is not surprising. Another choice can be found in Roussas (1991).
4. Assumptions (H4) and (H5) are technical.
5. Assumption (H8) is technical conditions imposed for the brevity of proofs.

By the definition of the conditional mode function we have

$$f^{x(1)}(\theta(x)) = 0.$$

Similarly, it follows that

$$\widehat{f}^{x(1)}(\widehat{\theta}_{kNN}(x)) = 0.$$

Furthermore, we assume that $f^{x(2)}(\theta(x)) \neq 0$, and $\widehat{f}^{x(2)}(\widehat{\theta}_{kNN}(x)) \neq 0$.

By a Taylor's expansion of $\widehat{f}^{x(1)}(\cdot)$ in the neighborhood of $\theta(x)$, we get we get

$$\widehat{f}^{x(1)}(\widehat{\theta}_{kNN}(x), u) = \widehat{f}^{x(1)}(\theta(x), u) + (\widehat{\theta}_{kNN}(x) - \theta(x)) \widehat{f}^{x(2)}(\theta^*(x), u),$$

with $\widehat{\theta}^*(x) \in (\widehat{\theta}_{kNN}(x), \theta(x))$. Because of the definition $\widehat{\theta}_{kNN}(x)$, we have

$$\widehat{\theta}_{kNN}(x) - \theta(x) = - \frac{\widehat{f}^{x(1)}(\theta(x), H_{n,k})}{\widehat{f}^{x(2)}(\widehat{\theta}^*(x), H_{n,k})}; \quad (5.13)$$

using equation (5.10), we can write

$$\widehat{\theta}_{kNN}(x) - \theta(x) = - \frac{\widehat{f}_N^{x(1)}(\theta(x), H_{n,k})}{\widehat{f}_N^{x(2)}(\widehat{\theta}^*(x), H_{n,k})}; \quad (5.14)$$

if the denominator does not vanish.

Finally, we will use the following notations :

$$a_j(x) = K^j(1) - \int_0^1 (K^j(s))' \zeta_0^x(s) ds, \text{ for } j = 1, 2.$$

Our main result is given in the following theorem.

Theorem 5.2.1. *Under assumptions (H1)-(H7) and (H8) for any $x \in \Upsilon$, we have*

$$\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))} \right)^{1/2} \left(\widehat{\theta}_{kNN}(x) - \theta(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \quad (5.15)$$

$\xrightarrow{\mathcal{D}}$ means the convergence in distribution, $\Upsilon = \{x : x \in \mathcal{S}, f^x(\theta(x)) \neq 0\}$ and

$$\sigma^2(x, \theta(x)) = \frac{a_2(x) f^x(\theta(x))}{[a_1(x) f^{x(2)}(\theta(x))]^2} \int_{\mathbb{R}} \left(G^{(2)}(t) \right)^2 dt. \quad (5.16)$$

Remark One can remark that the assumption imposed on the mixing coefficient $\alpha(k)$ under which Theorem 5.2.1 holds are assumptions (H3) and (H7.1).

Let $\alpha(k) = O(k^{-b})$ for some $b > 0$. In consequence, assumption (H7.1) holds if $b > \frac{2-2/\nu}{1-2/\nu}$.

For assumption (H7.1), we choose the small block $v_n = [(nh\phi_x(h))^{1/2}/\log n]$. Furthermore if $h = n^{-\zeta}$ and $\phi_x(h) = n^{-c}$, for some $\zeta > 0$ and $c > 0$. Theorem 5.2.1 holds if $0 < c < 1/2$, $\frac{1-2c}{6} < \zeta < \frac{1-2c}{5}$ and $b > \max \left\{ \frac{2+c}{4/3-c}, \frac{2-2/\nu}{1-2/\nu} \right\}$.

The detailed proof of the Theorem 5.2.1 is postponed to appendix.

5.3 Application

5.3.1 Application to prediction

For $n \in \mathbb{N}^*$, let $X_i(t), i = 1, \dots, n$ be functional random variables with $t \in \mathbb{R}$. For each curve $X_i(t)$, we have also at hand the value of some real response variables Y_i which corresponds to some modality of our problem.

The aim important question which is posed in force in practice is the following one : given a new curve $X_{n+1}(t) = x_{new}$, can we predict the corresponding response y_{new} ?.

This is a typical prediction problem for infinite-dimensional explanatory random variables.

The predictor estimator is defined by computing the quantity

$$\widehat{y}_{new} = \widehat{\theta}_{kNN}(x_{new}) = \arg \max_y \widehat{f}_{kNN}^{x_{new}}(y).$$

The theoretical predictor of x_{new} is clearly defined by equation (5.11) that is $y_{new} = \theta(x_{new})$.

Applying Theorem 5.2.1, we have corollary 5.3.1.

Corollary 5.3.1. *Under the hypotheses of Theorem 5.2.1 we have*

$$\left(\frac{h_G^3 k}{\sigma^2(x_{new}, \theta(x_{new}))} \right)^{1/2} \left(\widehat{\theta}_{kNN}(x_{new}) - \theta(x_{new}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \quad (5.17)$$

5.3.2 Confidence bands

Our main application of the above Theorem is the to build confidence interval for the true value of θ given curve $X = x$. A plug-in estimate for the asymptotic standard deviation $\sigma^2(x, \theta(x))$ can be obtained using the estimators $\widehat{\theta}_{kNN}(x)$, $\widehat{f}^{(x)}(\widehat{\theta}_{kNN}(x))$, $\widehat{f}^{(2)}(\widehat{\theta}_{kNN}(x))$ and $\widehat{a}_l(x)$ ($l = 1, 2$) of $\theta(x)$, $f^{(\cdot)}(\cdot)$ and $f^{(2)}(\cdot)$, $a_l(x)$, ($l = 1, 2$) respectively, that is

$$\widehat{\sigma}^2(x, \widehat{\theta}_{kNN}(x)) = \frac{\widehat{a}_2(x) \widehat{f}^{(x)}(\widehat{\theta}_{kNN}(x))}{\left[\widehat{a}_1(x) \widehat{f}^{(2)}(\widehat{\theta}_{kNN}(x)) \right]^2} \int_{\mathbb{R}} \left(G^{(2)}(t) \right)^2 dt. \quad (5.18)$$

Then $\widehat{\sigma}(x, \widehat{\theta}_{kNN}(x))$ can be used to get the following approximate $(1 - \zeta)$ confidence interval for $\theta(x)$

$$\widehat{\theta}_{kNN}(x) \pm \mu_{1-\zeta/2} \times \left(\frac{\widehat{\sigma}^2(x, \widehat{\theta}_{kNN}(x))}{h_G^3 k} \right)^{1/2}$$

where $\mu_{1-\zeta/2}$ denotes the $1 - \zeta/2$ quantile of the standard normal distribution.

We estimate empirically $a_l(x)$ by

$$\widehat{a}_l(x) = \frac{1}{h_G k} \sum_{i=1}^n K^l \left(\frac{d(x, X_i)}{\phi^{-1}(k/n)} \right).$$

This estimation is justified by the fact that, under (H1) and (H6), we have, (see Ferraty and Vieu (2006, p. 44)

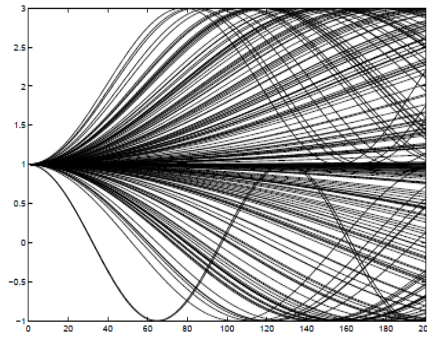
$$\frac{1}{h_G k} \mathbb{E}[K_1^l] \rightarrow a_l(x), \quad l = 1, 2.$$

Finally, the approximate $(1 - \zeta/2)$ confidence band, for any $x \in \Upsilon$, is

$$[\varrho_-(x), \varrho_+(x)] \text{ where } \varrho_{\pm}(x) = \widehat{\theta}_{kNN}(x) \pm \mu_{1-\zeta/2} \times \left(\frac{h_G^2 \sum_{i=1}^n K_i^2 \widehat{f}^{(x)}(\widehat{\theta}_{kNN}(x))}{\left(\sum_{i=1}^n K_i \widehat{f}^{(2)}(\widehat{\theta}_{kNN}(x)) \right)^2} \right)^{1/2}.$$

5.4 Simulation study

In this section we will show the effectiveness of k -NN method compared to the kernel estimation using simulated data. For this we considered a sample of a diffusion process on interval $[0, 1]$, $Z_1(t) = 2 - \cos(\pi t W)$ and $Z_2(t) = \cos(\pi t W)$, where W is the standard normal distribution and take $X(t) = A Z_1(t) + (1 - A) Z_2(t)$, where A is random variable Bernoulli distributed. We carried out the simulation with 200-sample of the curve X which is represented by the following graph :


 FIGURE 5.1 – The 200 curves X .

For the scalar response variable, we took $Y = Ar_1(X) + (1 - A)r_2(X) + \varepsilon$ where r_1 (resp. r_2) is the nonlinear regression model $r_1(X) = 0.25 \times \int_0^1 (X'(t))^2 dt + \varepsilon$, with ε is $U([0, 0.5])$ (resp. $r_2(X)$ is the null function) and ε is an α -mixing process generated by this model :

$$\varepsilon_i = \frac{1}{\sqrt{2}}(\varepsilon_{i-1} + \eta_i), \quad i = 1, \dots, 200,$$

with η_i is normally independent identically distributed random variables. We generated standard normal distributed random variable ε_0 independently of η_i .

We choose a quadratic kernel K defined by :

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{[0,1]}.$$

and the distribution function $G(\cdot)$ is defined by

$$G(x) = \int_{-\infty}^x \frac{15}{4}t^2(1 - t^2)\mathbb{I}_{[-1,1]}(t)dt. \quad (5.19)$$

In practice, the semi-metric choice is based on the regularity of the curves X . For this we use the semi-metric defined by L_2 -distance between the q^{th} derivatives of the curves. In order to evaluate the MSE (Mean Square Error) we proceed by the following algorithm :

- *Step 1.* We split our data into two subsets ; the first sample, of size $n=120$ corresponds to the learning sample which will be used, as a sample, to compute our conditional mode function estimators for the 80 remaining curves (considered as the test sample).
 - $(X_j, Y_j)_{j \in J}$ learning sample,
 - $(X_i, Y_i)_{i \in I}$ test sample.
- *Step 2.* We use the learning sample for computing the hazard function estimator $\hat{\theta}_{X_j}$, for all $j \in J$.
- *Step 3.* For each X_i in the test sample, we set : $i_* = \arg \min_{j \in J} d(X_i, X_j)$.

– *Step 4.* We put : $\forall i \in I$ and we take

$$\widehat{Y}_i = \widehat{\theta}(X_{i^*}),$$

where X_{i^*} is the nearest curve to X_i in the training sample.

And

$$\widetilde{Y}_i = \widehat{\theta}_{k_{opt}}(X_{k_{opt}}),$$

whereas the optimal number of neighbours k_{opt} is defined by

$$k_{opt} = \arg \min_k CV(k),$$

where

$$CV(k) = \sum_{i=121}^{200} (Y_i - \widehat{\theta}_{k_{NN}}^{-i}(x))^2,$$

with

$$\widehat{\theta}_{k_{NN}}^{-i}(x) = \arg \max_y (\widehat{f}_{k_{NN}}^x)^{-i}(y),$$

and

$$(\widehat{f}_{k_{NN}}^x)^{-i}(y) = \frac{h_G^{-1} \sum_{l=1, l \neq i}^{120} K(H_{n,k}^{-1}(x)d(x, X_l))G^{(1)}(h_G^{-1}(y - Y_l))}{\sum_{l=1, j \neq i}^{120} K(H_{n,k}^{-1}(x)d(x, X_l))}.$$

The error used to evaluate this comparison is the mean square error (MSE) expressed by

$$\frac{1}{\text{card}(I)} \sum_{i \in I} |Y_i - \widehat{T}(X_i)|^2,$$

where \widehat{T} designate the estimator used : kernel or k -NN method to estimate the conditional mode. Consequently, the k -NN method gives better results than the kernel method. This is confirmed by the $\text{MSE-}k\text{-NN} = 0.00101$ and $\text{MSE-Kernel} = 0.00652$.

– *Step 5.* We implement now our methodology and see how good the asymptotic normality in the determination of the confidence bands.

We define the confidence bands for all $i \in I$ by :

$$\left[\widehat{\theta}(X_{i^*}) + t_{0.975} \times \left(\frac{\widehat{\sigma}^2(X_{i^*}, \widehat{\theta}(X_{i^*}))}{h_G^3 k} \right), \widehat{\theta}(X_{i^*}) - t_{0.975} \times \left(\frac{\widehat{\sigma}^2(X_{i^*}, \widehat{\theta}(X_{i^*}))}{h_G^3 k} \right) \right]$$

where $t_{0.975}$ is the 5% quantile of a standard normal distribution.

We present our results by plotting the extremities of the predicted values versus the true values in the Figure 5.2.

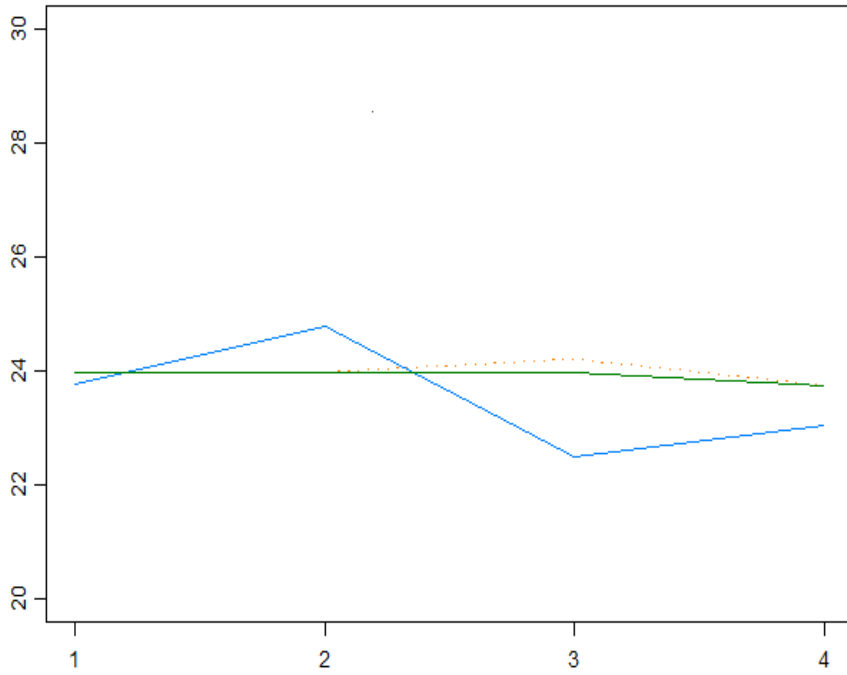


FIGURE 5.2 – The 95% conditional predictive bands.

5.5 Appendix

Using equation (5.14), with $D_n = H_{n,k}$, $c_n(D_n) = \widehat{f}_N^x(y, h)$, $c = f^x(y)$ and let $\lambda \in]0, 1[$, we choose D_n^- and D_n^+ as

$$W_x(D_n^-) = \sqrt{\lambda} \frac{k}{n}, \quad W_x(D_n^+) = \frac{1}{\sqrt{\lambda}} \frac{k}{n},$$

we have

$$\widehat{\theta}_{kNN}(x) - \theta(x) = \frac{-c_n^{(1)}(\theta(x), H_{n,k})}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})}. \quad (5.20)$$

Then, we obtain this decomposition :

$$\begin{aligned}
 \left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} (\widehat{\theta}_{kNN}(x) - \theta(x)) &= \left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left(\frac{-c_n^{(1)}(\theta(x), H_{n,k})}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})}\right) \\
 &= \left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left[\frac{-c_n^{(1)}(\theta(x), H_{n,k}) + c_n^{(1)}(\theta(x), D_n^+) - c_n^{(1)}(\theta(x), D_n^+)}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})}\right] \\
 &= -\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \frac{c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})} \\
 &+ \left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \frac{-c_n^{(1)}(\theta(x), H_{n,k}) + c_n^{(1)}(\theta(x), D_n^+)}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})} \\
 &- \left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \frac{\mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]}{c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k})}.
 \end{aligned} \tag{5.21}$$

Then, to state asymptotic normality, we show that the numerator of the first term of the right hand side of (5.21) suitably normalized is asymptotically normally distributed, the numerator of the last term goes to zero in probability as n goes to infinity and the denominator converges in probability to $(a_1(x) f^{x(2)}(\theta(x)))$.

For this, we remind that, under the same assumptions as Theorem 4.1, Ezzahrioui et al (2007) proved that

$$\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left(c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]\right) \xrightarrow{D} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \tag{5.22}$$

Lemma 4.11 of Ezzahrioui et al (2007) gives that

$$\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left(\mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]\right) \rightarrow 0, \text{ as } n \rightarrow \infty, \tag{5.23}$$

In the other hand, by hypothesis (H4), and the fact that $\mathbb{I}_{\{D_n^- \leq H_{n,k} \leq D_n^+\}} \xrightarrow{a.co.} 1$ when $\frac{k}{n} \rightarrow 0$ (see Burba et al. 2009), we have

$$c_n^{(1)}(\theta(x), D_n^+) \leq c_n^{(1)}(\theta(x), H_{n,k}) \leq c_n^{(1)}(\theta(x), D_n^-).$$

Using that

$$\begin{aligned}
 \left|c_n^{(1)}(\theta(x), H_{k,n}) - c_n^{(1)}(\theta(x), D_n^+)\right| &\leq \left|c_n^{(1)}(\theta(x), D_n^-) - c_n^{(1)}(\theta(x), D_n^+)\right| \\
 &\leq \left|c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]\right| \\
 &+ \left|\mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)]\right| \\
 &+ \left|\mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] - c_n^{(1)}(\theta(x), D_n^-)\right|.
 \end{aligned} \tag{5.24}$$

The equation

$$\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left|c_n^{(1)}(\theta(x), D_n^+) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)]\right| \rightarrow 0 \text{ as } n \rightarrow \infty, \tag{5.25}$$

is the same as (5.22) and Lemma 4.5 of Ezzahrioui et al (2007) gives that

$$\left(\frac{h_G^3 k}{\sigma^2(x, \theta(x))}\right)^{1/2} \left|c_n^{(1)}(\theta(x), D_n^-) - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)]\right| \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{5.26}$$

For the third term, we have :

$$\left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] - \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] \right| \leq \left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^+)] - c \right| + \left| \mathbb{E}[c_n^{(1)}(\theta(x), D_n^-)] - c \right|. \quad (5.27)$$

The almost complete convergence to 0 of these two terms is verified in (5.25) and (5.26).

On the other hand, to establish the convergence in probability of denominator in (5.21), note that

$$\left| c_n^{(2)}(\widehat{\theta}^*(x), H_{n,k}) - a_1(x)f^{x(2)}(\theta(x)) \right| \leq \left| c_n^{(2)}(\widehat{\theta}^*(x), D_n^+) - c_n^{(2)}(\theta(x), D_n^+) \right| + \left| c_n^{(2)}(\theta(x), D_n^+) - a_1(x)f^{x(2)}(\theta(x)) \right|. \quad (5.28)$$

By Lemma 4.5 in Ezzahrioui et al (2007) and the continuity of $c_n^{(2)}(., .)$ we deduce that the first term of (5.28) converges in probability to 0.

However, the limit of second term is consequence of the following inequality and the Lemma 4.3 for $j = 2$, Lemma 4.5 and proposition 4.2 of Ezzahrioui et al (2007)

$$\left| c_n^{(2)}(\theta(x), H_{n,k}) - a_1(x)f^{x(2)}(\theta(x)) \right| \leq \left| c_n^{(2)}(\theta(x), D_n^+) - a_1(x)f^{x(2)}(\theta(x)) \right|.$$

■

Références

1. Attouch, M., Bouabça, W. (2013). The k -nearest neighbors estimation of the conditional mode for functional data, *Rev. Roumaine Math. Pures Appl*, **58**, No. 4, pp. 393-415.
2. Attouch, M., Benchikh, T. (2012). Asymptotic distribution of robust k -nearest neighbour estimator for functional nonparametric models, *Mathematic Vesnic*, **64**, No. 4, pp. 275-285.
3. Abadie, A., and Imbens, G.(2006). Large Sample Properties of Matching Estimators for Average Treatment Effects, *Econometrica*, **74(1)**, 235-267.
4. Boente, G., and Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables, *Journal of Multivariate Analysis*, **25**, 90-99.
5. Boente, G. and Fraiman, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes, *Annals of Statistics*, **18**, 891-906.
6. Burba, F., Ferraty, F., Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression, *J. Nonparametric. Stat*, **21(4)**, 453-469.
7. Collomb, G. (1985). Nonparametric regression : an up to date bibliography. *Statistics*, **16**, 309-324.
8. Collomb, G., Härdle, W., Hassani, S. (1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf.* **15**, 227-236.
9. Delgado, M. (1992). A Semiparametric Generalized Least Squares in the Multivariate Nonlinear Regression Model, *Econometric Theory*, **8**, 203-222.
10. Delgado, M. A., and Stengos, T. (1994). Semiparametric Testing in Non-nested Econometric Models, *Review of Economic Studies*, **61(2)**, 291-303.
11. Delsol, L. (2008b). Advances on asymptotic normality in nonparametric functional Time Series Analysis *Statistics*.
12. Ezzahrioui, M., Ould-Said, E. (2007). Prévission dans les modèles conditionels en dimension finie, PhD Thesis.
13. Ferraty, F., and Vieu, P. (2006). Nonparametric Functional Data Analysis : Theory and practice, *Springer series in statistics*, Springer. New York.
14. Ferraty, F., Laksaci, A., Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Processes*, **9**, 47-76.
15. Ferraty, F., Mas, A., Vieu, P. (2007) .Advances in nonparametric regression for functional variables. *Aust. New Zealand. J. of Satist*, **49**, 1-20.
16. Györfi, L., Härdle, W., Sarda, P., Vieu, P. (1989). Nonparametric curve estimation from time series, *Lecture Notes in Statistics*, Springer-Verlag, **60**.
17. Györfi, L., Kohler, M., Krzyzak, A., and Walk, H. (2002). A distribution-free theory of nonparametric regression. *Springer Series in Statistics*.

18. Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching as an Econometric Evaluation Estimator, *Review of Economic Studies*, **65**(2), 261-94.
19. Jacho-Chávez, D. T. (2008). k nearest-neighbor estimation of inverse density weighted expectations, *Economics Bulletin*, **3**(48), 1-6.
20. Jun, S. J., and Pinkse, J. (2009a). Efficient Semiparametric Seemingly Unrelated Quantile Regression Estimation, *Econometric Theory*, **25**(05), 1392-1414.
21. Jun, S. J., and Pinkse, J. (2009b). Semiparametric tests of conditional moment restrictions under weak or partial identification, *Journal of Econometrics*, **152**(1), 3-18.
22. Jun, S. J., and Pinkse, J. (2009c). Testing under weak identification with conditional moment conditions, Unpublished Manuscript.
23. Liu, Z., and Lu, X. (1997). Root-n-consistent Semiparametric Estimation of Partially Linear Models Based on k-NN Method, *Econometric Reviews*, **16**(4), 411-420.
24. Li, Q. (2006). Testing Parametric Regression Functional Forms Using k-Nearest Neighbor Method, Unpublished manuscript.
25. Louani, D. (1998). Large deviations limit theorems for the kernel density estimator, *Scand. J. Statist.*, **25**(1) :243-253.
26. Louani, D., Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis, *J. Nonparametric Statist*, **11**, No.4, 413-442.
27. Masry, E. (2005). Nonparametric regression estimation for dependent functional data, Asymptotic normality. *Stoch. Proc. and their Appl*, **115**, 155-177.
28. Mehra, K.L., Ramakrishnaiah, Y.S., and Sashikala,P. (2000). Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode. *Ann. Inst. Statist. Math.* **52**, 630-645.
29. Newey, W. K. (1990). Efficient Instrumental Variables Estimation of Nonlinear Models, *Econometrica*, **58**(4), 809-37.
30. Ould-Saïd, E. (1993). Estimation nonparamétrique du mode conditionnel. Application à la prévision. *C. R. Acad. Sci. Paris, Série I*, **316**, 943-947.
31. Ould-Saïd, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Statist*, **24**, 231-239.
32. Pinkse, J. (2006). Heteroskedasticity Correction and Dimension Reduction, Unpublished Manuscript.
33. Robinson, P. M. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form, *Econometrica*, **55**(4), 875-891.
34. Robinson, P. M. (1995). Nearest-neighbour estimation of semiparametric regression models, *Journal of Nonparametric Statistics*,**5**(1), 33-41.

35. Roussas, G.G. (1991). Estimation of a transition distribution function and its quantiles in Markov processes : Strong consistency and asymptotic normality, *Nonparametric Functional Estimation and Related Topics*, *Kluwer Academic Publishers, Dordrecht*, 443-362.
36. Romano, On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist*, **16**, 629-647.
37. Samanta, M., and Thavaneswaran, A. (1990). Non-parametric estimation of conditional mode. *Comm. Statist. Theory and Meth.*, **16**, 4515-4524.
38. Tran, L. T., and Yakowitz, S. (1993). Nearest neighbor estimators for random fields, *Journal of Multivariate Analysis*, **44**, 23-46.
39. Vieu, P. (1996). Estimation non paramétrique d'une densité multidimensionnelle : choix optimaux d'estimateurs. *C. R. Acad. Sci. Paris, 322, Série I*, 395-398.
40. Yakowitz, S. (1987). Nearest-neighbor methods for time series analysis, *Journal of Time Series Analysis*, **8(2)**, 235-247.

Chapitre 6

Annexe 1 : Commentaires et perspectives

L'objectif de ce chapitre est de de mettre l'accent sur les problèmes rencontrés.

6.1 Quelques questions pratiques

6.1.1 Choix de la semi-métrie

Il est clair que le choix de la semi-métrie joue un rôle déterminant notamment dans l'amélioration des vitesses de convergence en augmentant la concentration de la mesure. Un exemple explicite du choix de semi-métrie vérifiant la propriété indiquée est donné dans Ferraty et al (2002). Cependant, à l'heure actuelle, il n'existe aucune méthode automatique qui offre un choix optimal de la semi-métrie. Ferraty et Vieu (2004) ont considéré une famille de semi-métries, définie par :

$$\forall (x_1, x_2) \in \mathcal{F} \times \mathcal{F}, d_q^{deriv}(x_1, x_2) = \left(\int (x_1^{(q)}(t) - x_2^{(q)}(t))^2 dt \right)^{1/2}, q \in \mathbb{N},$$

où $x^{(q)}$ est la dérivée d'ordre q . En utilisant des données spectrométriques, les auteurs ont montré que la métrie qui minimise l'erreur de prédiction pour ces données est obtenue pour $q = 2$. Nous avons ainsi choisi d'utiliser cette distance dans nos applications.

Plusieurs familles de semi-métrie sont utilisées, citons par exemple, la semi-métrie de l'analyse en composantes principales fonctionnelles introduite par Besse et al (1997) et la semi-métrie de modèle d'indice fonctionnel utilisée par Ferraty et al (2002).

- **La semi-métrie de l'analyse en composantes principales fonctionnelles** : c'est une métrie introduite par (Besse et al (1997)). Elle est définie par la topologie de la norme L^2 dont le calcul est essentiellement basé sur l'analyse des composantes principales. Plus précisément,

on cherche les q premier vecteurs propres de l'opérateur de variance covariance empirique, associés au q plus grandes valeurs propres, la métrique est alors définie par :

$$\forall (x_1, x_2) \in \mathcal{F} \times \mathcal{F}, d_q^{deriv}(x_1, x_2) = \int (\prod_q (x_1^{(q)}(t) - x_2^{(q)}(t)))^2 dt, q \in \mathbb{N},$$

où \prod_q est le projecteur orthogonal sur l'espace L^2 engendré par les q vecteurs propres associés aux q plus grandes valeurs propres.

- **La semi-métrique de modèle d'indice fonctionnelle** : pour une variable fonctionnelle X dans un espace de Hilbert \mathcal{H} muni de produit scalaires $\langle \cdot, \cdot \rangle$, on définit la semi-métrique de modèle d'indice fonctionnel par

$$\forall \alpha, x, y \in \mathcal{H} \quad d_\alpha(x, y) = | \langle x - y, \alpha \rangle |.$$

Dans nos applications nous avons procédé par des techniques de type validation croisée pour sélectionner le paramètre $(p; q; \alpha)$ de la semi-métrique, bien sûr après avoir choisit la famille de semi-métriques. Le choix de ce dernier est motivé par la régularité des données étudiées. Autrement dit, si les courbes sont lissent on prend la famille de semi-métrique de la dérivée, par contre s'ils présentent quelques discontinuités la semi-métrique de l'ACP est la plus appropriée. Il est à noter que Ait Saïdi et al (2008) ont élaboré une méthode de sélection automatique du choix du paramètre α de la semi-métrique d'indice fonctionnel pour le cas de la régression. Cependant le sujet reste ouvert pour d'autres paramètres fonctionnels.

On peut retenir que le choix de la semi-métrique permet à la fois de prendre en compte des situations plus variées et de pouvoir contourner le fléau de la dimension. Ce choix ne doit cependant pas être fait à la légère mais en prenant en compte la nature des données et du problème étudié.

6.1.2 Choix du noyau

En générale on utilise le noyau gaussien en raison de sa simplicité et ses propriétés asymptotiques. Cependant l'inconvénient principal de ce noyau est qu'il attribue des poids positif pour des coordonnées qui sont à l'extérieur du support lorsque on cherche l'estimateur de la densité de probabilité d'une variable aléatoire bornée ou semi bornée. Cela cause le problème du biais aux bornes et donne un estimateur non consistant. La solution consiste à utiliser des noyaux asymétriques et adaptés qui n'assignent aucun poids à l'extérieur du support. Chen ((1999)et (2000)) propose respectivement le noyau Beta pour les densités à support compact et le noyau gamma pour les densités à variables à support positif (c'est-à-dire sur $[0; +1]$).

6.1.3 Choix du paramètre de lissage

Des différentes méthodes d'estimation non-paramétriques ont leurs propres avantages dans des situations différentes. Sans aucun doute la méthode k -NN pourrait être la méthode préférée dans

quelques applications, par exemple quand les points de données sont inégalement distribués dans son support, comme est souvent le cas pour des données non-expérimentales dans sciences sociales. Des méthodes à noyau avec un paramètre de lissage fixe (sur une gamme de données entière) peuvent contenir peu de points de données dans une certaine gamme du support et peuvent ainsi mener à l'estimation incertaine et des résultats d'inférence.

Dans ce cas un chercheur peut préférer d'utiliser la méthode k -NN qui utilise toujours (le plus proche) k points de données dans l'estimation non-paramétrique, ou une certaine méthode adaptative à noyau qui utilise des paramètres de lissage différents pour des différente gamme de points de données (voir Ruppert et al. (2003); Schucany (2004)).

Il est aussi bien établi que la sélection de paramètres de lissage a d'importance cruciale dans l'estimation non-paramétrique. Dans le cadre d'estimer une fonction de régression non-paramétrique, des méthodes axées sur les données diverses sont suggérées et étudiées.

Avec des données indépendantes ou faiblement dépendantes, la méthode "least-squares-leave-one-out cross validation" est la méthode la plus utilisée pour choisir le paramètre de lissage, voir Härdle et al. (1988, 1992), Gao et Tong (2004).

Li (1987) a montré que pour l'estimateur k -NN de régression sous homoskedasticity conditionnel, il est asymptotiquement optimal de choisir k par le CV Généralisé, ou le CV (défini ci-dessus).

Cependant Li (1987) démontre que la méthode de validation croisée "leave-one-out cross validation" peut choisir le k d'une façon optimale dans le sens de minimiser l'erreur carrée d'estimation intégrée.

Andrew (1991) a généralisé ce résultat au cas de heteroskedasticity et il a montré que le CV est asymptotiquement optimal.

Ouyang et al (2006) prolonge le résultat de Li (1987) et tire le taux de convergence de "leave-one-out cross validation" \hat{k}_{CV} à un certain point de référence optimal à valeurs non-stochastique de paramètre de lissage optimales.

Dans nos applications, nous avons procédé en choisissant le paramètre de lissage par la méthode de validation croisée sur le nombre de voisins les plus proches et qui a montré une compatibilité avec nos jeux de données. Bien qu'il n'existe pas une étude théorique, à notre connaissance montrons l'optimalité asymptotique de cette méthode dans le cadre fonctionnel.

Néanmoins, nous renvoyons à Rachdi et Vieu (2007), Benhenni et al. (2007) pour le choix optimal du paramètre de lissage dans la régression classique et Madani et al. (2008) pour le cas de la densité conditionnelle.

6.1.4 A propos de l'existence de la version régulière de la probabilité conditionnelle

Le mode conditionnel exprimée en fonction d'une densité conditionnelle de Y sachant X est un paramètre conditionnel. Autrement dit, ce modèle non paramétrique est implicitement lié à la dis-

tribution conditionnelle de Y sachant X . En conditionnant par des variables aléatoires appartenant à des espaces métriques séparables.

Evidemment, dans toute la thèse on suppose que la version régulière de la probabilité conditionnelle existe dans un espace semi métrique. Cependant ce problème est encore ouvert en probabilité et la littérature sur le sujet n'a pas encore confirmé l'existence ou non de cette version dans ces structures topologiques assez générales. En conditionnant par des variables aléatoires appartenant à des espaces métriques séparables, le théorème de Jirina (Tortrat, 1971,P.77) a montré l'existence de la version régulière de cette probabilité conditionnelle.

Leão, Fragoso et Ruffino (1999, 2004) ont donnés des conditions nécessaires et suffisantes pour l'existence de cette version régulière sur des structures topologiques moins restrictives que ceux proposé par Jirina (1959) en employant les notions de complétude, compacité, et séparabilité. Mais la question reste ouverte pour des espace semi métrique généraux.

En terme d'applications, les considérations topologiques de Jirina pour l'existence de cette version reste faible par rapport à l'énorme potentialité qu'offre la modélisation par des données fonctionnelles dans des structures topologiques assez générales. Dans notre thèse nous avons opté pour une présentation sous la structure semi-métrique en supposant l'existence de cette version régulière pour quelque raison. La première est liée aux nombreuses applications offertes en considérant un espace semi métrique au lieu d'un espace métrique qui permet à la fois de prendre en considération des situations plus variées et de contourner le problème du fléau de la dimension ce qui permet d'avoir une grand souplesse dans le choix d'indices de proximité entre des objets fonctionnels. La deuxième vient du fait que les semi-métriques utilisées sur les variables initiales sont équivalentes à des métrique agissant sur les variables transformées, situation dans la quelle l'existence de la version régulière est vérifiée dès que l'espace métrique considéré est complet et séparable.

Finalement nous signalons que dans d'autres contextes statistiques fonctionnelle comme la théorie des problèmes inverse, la plupart des auteurs qui ont étudié les modèles non paramétriques conditionnels (voir par exemple Ferraty et Vieu (2006a), Laksaci (2005), Lihtinen et al. (1989),...) Supposent implicitement l'existence de cette version régulière dans les espaces non nécessairement métriques.

6.2 Perspectives

6.2.1 Questions ouvertes

Le travail que nous avons développé dans cette thèse offre de nombreuses perspectives.

Parmi lesquelles :

1. **Le choix de la semi-métrique.** Nous avons vu qu'un choix judicieux de la semi-métrique contribue à l'optimalité de l'estimateur, de ce faite, trouver une famille de semi-métrique "adaptatives" (par exemple celle déduite du modèle d'indice fonctionnel) améliorera la performance de prédiction de nos estimateurs.

2. **La convergence en norme L^p .** En faisant une adaptation des outils développés par Niang et Rhomari (2003) on pourrait étudier la convergence en norme L^p .
3. **Le choix du paramètre de lissage.** Le choix de ce paramètre est très important, car il permet l'amélioration de la qualité de l'estimateur. Une généralisation du travail de Ouyang et al (2006) permettra peut être de traiter ce problème.

D'autres thèmes peuvent être abordés à long terme tels le conditionnement par p variables fonctionnelle une combinaison linéaire des ces p variables par la méthode des plus proches voisins, \dots .

Chapitre 7

Annexe 2 : Définitions et outils probabilistes

Dans ce chapitre, nous allons rappeler quelques définitions et théorèmes utilisés le long de cette thèse.

7.1 Notations et définitions

Définition 1.1.1 (Fonction noyau)

Soit E un espace de \mathbb{R}^p (où $1 \leq p \leq +\infty$) et soit K une fonction définie sur E .

Si K est une fonction positive, décroissante, symétrique et normalisée (c.à.d $\int_E K(u)du = 1$), nous disons que K est une fonction noyau (Kernel).

Définition 1.1.2 (Noyau d'ordre α)

Soit E un espace de \mathbb{R}^p (où $1 \leq p \leq +\infty$) et soit K une fonction noyau définie sur E .

Nous disons que K est un noyau d'ordre α (avec $\alpha \in \mathbb{N}^*$), lorsque :

$$\forall i_1, i_2, \dots, i_p \in \mathbb{N}^* : (\forall j \in \{1, 2, \dots, p\} : i_j < \alpha) \Rightarrow \int_E u_1^{i_1} u_2^{i_2} \dots u_p^{i_p} K(u_1, u_2, \dots, u_p) du_1 du_2 \dots du_p = 0.$$

Définition 1.1.3 (Convergence presque complète)

Nous disons que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge presque complètement vers une variable aléatoire réelle X , si et seulement si :

$$\forall \epsilon > 0, \quad \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \epsilon) < \infty,$$

et la convergence presque complète de $(X_n)_{n \in \mathbb{N}}$ vers X est notée par :

$$X_n \xrightarrow[n \rightarrow +\infty]{p.co.} X.$$

Si $X_n \xrightarrow[n \rightarrow +\infty]{p.co.} X$, et $Y_n \xrightarrow[n \rightarrow +\infty]{p.co.} Y$ alors :

- a) $\frac{1}{X_n} \xrightarrow[n \rightarrow +\infty]{p.co.} \frac{1}{X}$.
- b) $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{p.co.} X + Y$.
- c) $X_n Y_n \xrightarrow[n \rightarrow +\infty]{p.co.} X Y$.

Définition 1.1.4 (Vitesse de convergence presque complète)

Nous disons que la vitesse de convergence presque complète de la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ vers X est d'ordre u_n si et seulement si :

$$\exists \epsilon_0 > 0, \quad \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \epsilon_0 u_n) < \infty,$$

et nous notons :

$$X_n - X = O_{p.co.}(u_n).$$

Si $X_n - X = O_{p.co.}(u_n)$, et $Y_n - Y = O_{p.co.}(u_n)$, avec $\lim_{n \rightarrow +\infty} u_n = 0$, alors :

- a) $\frac{1}{X_n} - \frac{1}{X} = O_{p.co.}(u_n)$.
- b) $(X_n + Y_n) - (X + Y) = O_{p.co.}(u_n)$.
- c) $(X_n Y_n) - (X Y) = O_{p.co.}(u_n)$.

Si $T_n = O_{p.co.}(u_n)$, et $S_n \xrightarrow[n \rightarrow +\infty]{p.co.} S$, alors :

- a) $T_n S_n = O_{p.co.}(u_n)$.
- b) $\frac{T_n}{S_n} = O_{p.co.}(u_n)$.

Définition 1.1.5 (Convergence en distribution)

La convergence en distribution de suites de variables aléatoires est un concept appartenant plus spécifiquement à la théorie des probabilités, utilisé notamment en statistique et dans l'étude des processus stochastiques. La convergence en distribution est souvent notée en ajoutant la lettre \mathcal{D} (ou \mathcal{L} en loi) au-dessus de la flèche de convergence.

La suite $(X_n)_{n \in \mathbb{N}}$ converge en distribution (ou en loi) vers X si

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{ou bien} \quad X_n \xrightarrow{\mathcal{L}} X$$

7.2 Outils de probabilité utilisés

Cas I.I.D. :

Soit $Z_1; \dots; Z_n$ une suite de variables aléatoires indépendantes identiquement distribuées.

Corollaire 1.2.1 $\forall m \geq 2, \exists c_m > 2 / \mathbb{E} |Z_1^n| \leq c_m a^{2(m-1)},$

alors :

$$\forall \epsilon > 0, \quad \mathbb{P} \left(\left| \sum_{i=1}^n Z_i - \mathbb{E}[Z_i] \right| > \epsilon n \right) \leq 2 \exp \left(- \frac{\epsilon^2 n}{2 a^2 (1 + \epsilon)} \right).$$

Supposons que Z_i dépend de n , ($Z_i = Z_{i;n}$), et si

$$\forall m \geq 2, \exists c_m > 2 / \mathbb{E} |Z_1^n| \leq c_m a^{2(m-1)},$$

et si la suite $(u_n = n^{-1} a_n^2 \log n)_{n \in \mathbb{N}}$ est telle que : $\lim_{n \rightarrow +\infty} u_n = 0$, alors :

$$\frac{1}{n} \sum_{i=1}^n Z_i = O(\sqrt{u_n}) \text{ p.co.}$$

Corollaire 1.2.2 Si $\exists M > 0$, tel que $Z_1 \leq M$, et notons $\sigma^2 = \text{var}(Z_1)$, on a

$$\forall \epsilon > 0, \quad \mathbb{P} \left(\left| \sum_{i=1}^n Z_i - \mathbb{E}[Z_i] \right| > \epsilon n \right) \leq 2 \exp \left(- \frac{\epsilon^2 n}{2 \sigma^2 (1 + \epsilon \frac{M}{\sigma^2})} \right).$$

Supposons que Z_i dépend de n , ($Z_i = Z_{i;n}$), si $\exists M = M_n < 1$, tel que $|Z_1| \leq M$, si la suite $(u_n = n^{-1} a_n^2 \log n)_{n \in \mathbb{N}}$ est telle que : $\lim_{n \rightarrow +\infty} u_n = 0$, et si $\frac{M}{\sigma^2} < C < \infty$, alors

$$\frac{1}{n} \sum_{i=1}^n Z_i = O(\sqrt{u_n}) \text{ p.co.}$$

Cas du mélange :

Proposition 7.2.1. *Supposons que $(T_n)_{n \in \mathbb{Z}}$ est α -mélangeantes. Notons pour $k \in \mathbb{Z}$, la variable aléatoire réelle \mathcal{T} (resp. \mathcal{T}') qui est*

$\sigma(T_i; -\infty < i \leq k)$ -mesurable (resp. $\sigma(T_i; n+k \leq i \leq +\infty)$ -mesurable).

1. *Si \mathcal{T} et \mathcal{T}' sont bornées, alors*

$$\exists C > 0, \text{cov}(\mathcal{T}; \mathcal{T}') \leq C \alpha(n)$$

2. *Si pour quelques nombres réels positive p , q et r tel que $p^{-1} + q^{-1} + r^{-1} = 1$, on a $\mathbb{E}[\mathcal{T}^p] < 1$ et $\mathbb{E}[\mathcal{T}'^q] < 1$*

$$\exists C > 0, \text{cov}(\mathcal{T}; \mathcal{T}') \leq C \mathbb{E}[\mathcal{T}^p]^{1/p} \mathbb{E}[\mathcal{T}'^q]^{1/q} \alpha(n)^{1/r}.$$

Notons $s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{cov}(W_i; W_j)|$, où $(W_n)_{n \in \mathbb{Z}}$ est une suite de variables aléatoires identiquement distribuées arithmétiquement α -mélangeantes avec un taux $a > 1$.

Proposition 7.2.2. 1. S'il existe $p > 2$ et $M > 0$ tel que $\forall t > M; \mathbb{P}(|W_1| > t) \leq t^{-p}$, alors on a pour tout $r \geq 1$, et $\epsilon > 0$, pour quelque $C < \infty$:

$$\mathbb{P} \left(\left| \sum_{i=1}^n W_i \right| > \epsilon \right) \leq C \left[\left(1 + \frac{\epsilon^2}{r s_n^2} \right)^{-r/2} + n r^{-1} \left(\frac{r}{\epsilon} \right)^{a+1} p / (a+p) \right].$$

2. S'il existe $M < 1$ tel que $|W_1| \leq M$, alors on a pour tout $r \geq 1$, et pour quelque $C < 1$:

$$\mathbb{P} \left(\left| \sum_{i=1}^n W_i \right| > \epsilon \right) \leq C \left[\left(1 + \frac{\epsilon^2}{r s_n^2} \right)^{-r/2} + n r^{-1} \left(\frac{r}{\epsilon} \right)^{a+1} \right].$$

Inégalité C_r

$$\mathbb{E}[|X + Y|^r] \leq c_r \mathbb{E}[|X|^r] + c_r \mathbb{E}[|Y|^r]$$

où $c^r = 1$ (resp. 2^{r-1}) selon que $r \leq 1$ (resp. $r \geq 1$).

La méthode de validation croisée

La validation croisée est une méthode très populaire depuis 40 ans au sein des praticiens pour estimer la performance d'une méthode. cette méthode est utilisée dans la résolution d'un grand nombre de problèmes : la sélection de modèle, l'adaptivité et l'identification. Elle peut être vue comme une application particulière des méthodes dites de rééchantillonnage.

En pratique, les jeux de données sont rarement suffisamment grands pour que l'erreur de test ainsi calculée estime correctement l'erreur de prédiction du modèle. La validation croisée est une alternative très populaire pour gérer la parcimonie des données. Il s'agit de découper le jeu de données en K groupes tirés aléatoirement qui font successivement servir d'ensemble de test. On peut alors calculer une erreur de test pour chacun des groupes et en faire la moyenne, ce qui constitue l'estimateur de l'erreur de test par validation croisée. Notons $CV(\mathcal{G})$ cette erreur pour un modèle \mathcal{G} donné.

On introduit également une fonction $\kappa : \{1; \dots, n\} \rightarrow \{1; \dots, K\}$ qui indique dans quel groupe de validation se trouve l'observation i . Ainsi, $\kappa(i) = k$ si l'observation i se trouve dans le groupe k de validation. Si on note $\widehat{\mathcal{G}}^{-k}$ le modèle estimé sur l'ensemble des données privé du k^e groupe, alors l'erreur de validation croisée s'écrit

$$CV(\mathcal{G}) = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{\mathcal{G}}^{-\kappa(i)}(X))^2. \quad (7.1)$$

Ainsi, y_i , la réponse observée, est à comparer avec $\widehat{\mathcal{G}}^{-\kappa(i)} = \widehat{y}_i^{-\kappa(i)}$, la valeur prédite du modèle estimé sur les données privés de l'ensemble de test où se trouve l'observation i . Lorsque $K = n$, on retrouve le cas extrême correspondant au leave-one-out.

En fait, il y a au moins trois techniques de validation croisée : "testset validation" ou "holdout method", " k -fold cross-validation" et "leave-one-out cross-validation (LOOCV)".

– La première méthode est très simple, il suffit de diviser l'échantillon de taille n en échantillon d'apprentissage ($> 60\%$ de l'échantillon) et échantillon de test. Le modèle est bâti sur l'échantillon

d'apprentissage et validé sur l'échantillon de test. L'erreur est estimée en calculant un test, une mesure ou un score de performance du modèle sur l'échantillon de test, par exemple l'erreur quadratique moyenne.

- Dans la seconde, on divise k fois l'échantillon, puis on sélectionne un des k échantillons comme ensemble de validation et les $(k - 1)$ autres échantillons constitueront l'ensemble d'apprentissage. On calcule comme dans la première méthode l'erreur quadratique moyenne. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les $(k - 1)$ échantillons qui n'ont pas encore été utilisés pour la validation du modèle. L'opération se répète ainsi k fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. La moyenne des k erreurs quadratiques moyennes est enfin calculée pour estimer l'erreur de prédiction.
- La troisième méthode est un cas particulier de la deuxième méthode où $k = n$, c'est-à-dire que l'on apprend sur $(n - 1)$ observations puis on valide le modèle sur la n ème observation et l'on répète cette opération n fois.

Références

1. Ait-Saïdi, A., Ferraty, F., Kassa, R., Vieu, P. (2008). Choix optimal du paramètre fonctionnel dans le modèle à indice fonctionnel simple. *C. R. Math. Acad. Sci., Paris*, **346**, No. 3-4, 217-220.
2. Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817-858.
3. Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist*, **22**, No. 3, 353-369.
4. Besse, P. et Ramsay, J.O. (1986). Principal components analysis of sampled functions. *Psychometrika*, **51 (2)** 285-311.
5. Besse, P., Cardot, H., Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Computational Statistics and Data Analysis*, **24**, 255-270.
6. Bosq, D. (2000). Linear Processes in Function Spaces, *Theory and Applications Lecture Notes in Statistics Springer-Verlag New York*, 149.
7. Chen, S. (1999). A Beta Kernel Estimation for Density Functions, *Computational Statistics and Data Analysis*, **31**, 131-145.
8. Chen, S. (2000). Probability Density Functions Estimation Using Gamma Kernels, *Annals of the Institute of Statistical Mathematics*, **52**, 471-480.
9. Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE TRANSACTIONS ON INFORMATION THEORY*, **13(1)** :21-27.
10. Crambes, C., Kneip, A. and Sarda, P. (2007). Smoothing splines estimators for functional linear regression submitted.
11. Craven, P., and Wahba, G. (1979). Smoothing noisy data with spline functions : Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, **31** : 377-403
12. Cuevas, A. et Fraiman, R. (2004). On the bootstrap methodology for functional data. *Proceedings in Computational Statistics Physica, Heidelberg*, 127-135
13. Dabo-Niang, S. and Rhomari, N. (2003). Kernel regression estimation when the regressor takes values in metric space. *C. R. Math. Acad. Sci. Paris*, **336 (1)** 75-80.
14. Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference *J. Multivariate Anal.* **12 (1)** 136-154.
15. Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés, *Compte Rendus de l'Académie des Sciences Paris*, **330**, 403-406.

16. Ferraty, F. et Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist*, **17** (4), 545-564.
17. Ferraty, F., and Vieu, P. (2004). Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination. *J. Nonparametr. Stat*, **v16**. 11-127.
18. Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice . *Springer-Verlag*.
19. Fix, E., and Hodges, J. (1989). Discriminatory analysis. nonparametric discrimination : Consistency properties. *International Statistical Review*, **53(3)** :238-247.
20. Gao, J., and Tong , H. (2004). Semiparametric non-linear time series model. *Journal of Royal Statistical Society*, **66**, 321-36.
21. Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis J.R. *Stat. Soc, B* **68** (1) 109-126.
22. Härdle, W., Hall, P., and Marron, J.S. (1988). How far are automatically chosen regression smoothing parameters from their optimum ?, *Journal of American Statistical Association*, **83**, 86-99.
23. Härdle, W., Hall, P., and Marron, J.S. (1992). Regression smoothing parameters that are not far from their optimum. *Journal of American Statistical Association*, **87**, 227-33.
24. Jirina, M. (1959). On regular conditional probabilities. *Journal math. tchécoslovaque*, **9**, No. 84, 445-450.
25. Laksaci, A. (2005). Contribution aux modèles non paramétriques conditionnels pour variables explicatives fonctionnels. Thèse de doctorat, université de Toulouse.
26. Laksaci, A., Madani, M., Rachdi, M. (2008). Kernel conditional density estimation when the regressor is valued in semi-metric space : how to select the bandwidth.
27. Leão, D.Jr., Fragoso, M.D., Ruffino, P. (1999). Characterizations of Radon spaces. *Statist. Probab. Lett*, **42**, No.4, 409-413.
28. Leão, D.Jr., Fragoso, M.D., Ruffino, P. (2004). Regular conditional probability, desintegration of probability and Radon spaces. *Proyecciones*, **23**, No.1, 15-29.
29. Lehtinen, M.S., Paivarinta, L., Somersalo, E. (1989). Linear inverse problems for generalized random variables. *Invers. Probl*, **5**, No.4, 599-612.
30. Li, K. C. (1987). Asymptotic optimality for c_p , c_{cl} , cross-validation, and generalized cross-validation. *Annals of Statistics*, **15**, 958-75.
31. Li, Q., and Racine, J. S. (2004). Nonparametric Econometrics : Theory and Practice. Princeton University Press.
32. Liu, Z., and Lu, X. (1997). Root-n-consistent semiparametric estimation of partially linear models based on k -nn method. *Econometric Reviews*, **16(4)** :411-420.

33. Loftsgaarden, D., and Quesenberry, C. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, **36(3)** :1049-1051.
34. Mack, Y. L., and Rosenblatt, M. (1979). Multivariate k -nearest neighbor density estimates. *Journal of Multivariate Analysis*, **9(1)** :1-15.
35. Mallow, C.L. (1973). Some comments on C_p , *Technometrics*, **15** : 661-675.
36. Ouyang, D., Li, D., and LI, Q. (2006) . Cross-validation and non-parametric k nearestneighbour estimation. *Econometrics Journal*, **9** :448-471.
37. Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *J. Statist. Plann. Inference*, **137 (3)** 829-840.
38. Rachdi, M., Vieu, P.(2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *J. Statist. Plann. and Inference*, **137**, 2784-2801.
39. Ramsay, J. and Silverman, B. (1997). Functional Data Analysis, *Springer- Verlag, New York*.
40. Ramsay, J. and Silverman, B. (2005). Functional Data Analysis (Second Edition). *Springer-Verlag, New York*.
41. Rossi, F., Delannay, N., Conan-Guez, B. et Verleysen, M. (2005). Representation of Functional Data in Neural Networks *Neurocomputing*, **64** 183-210.
42. Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). Semiparametric regression. Cambridge Univ. Press. Cambridge, UK.
43. Schucany, W. R. (2004). Kernel smoothers : An overview of curve estimators for the first graduate course in nonparametric statistics source. *Statistical Science*, **19**, 663-675.
44. Stone, C.J. (1974). Cross-validatory choices and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, **36** : 111-147.
45. Yao, F. and Lee, T.C.M. (2006). Penalised spline models for functional principal component analysis J.R. *Stat. Soc. B*, **68 (1)** 3-25.